VOLUME IV
TEST MANAGEMENT PHASE

CHAPTER 4

LOGISTICS FLIGHT TESTING

APRIL 1991

USAF TEST PILOT SCHOOL
EDWARDS AFB CA

LOGISTICS TESTING

"Flying's the fun part. Keeping 'em flying's a bitch!" (Anon. loggie)

Why read this?

Why know this?

Why remember this?

Usually, the title above is enough to turn off any bloodthirsty fighter pilot worth his fangs, any bomber worth his nukes, any IP worth his students, and any trash hauler worth his TDY pay. But you golden arms and golden brains who will soon bid a fond (?) farewell to the "world-famous Test Pilot School" in the Mojave have got to "have a dime" in this type of testing as much as in all the other kinds. We'll boil a lot of this text down to the issues that you've gotta know when you get back into the Real World. The WHYs listed above are answered by the simple fact that the fighters, bombers, trainers and transports of tomorrow, along with their bombs, rockets, bullets, lasers, and particle beam transmorgifiers, are going to be tested and passed/failed by YOU. The good parts and the bad parts. The things you absolutely love and the things you didn't notice until the hardware gets out to the guys in the field. All of it will be under YOUR microscope. All of it will appear on YOUR flight card or on the flight cards of the folks who work for you. Since what the Washingtonians will pay for is based on YOUR reports on the performance, suitability, spec compliance, reliability, and supportability of this expensive, state-of-the-art equipment, it's somewhat important that you soon-to-be Test Directors, Squadron Commanders, and Test Managers understand the whole big picture.

Logistics is part of that BIIIIIIG PICTURE. A cosmic advanced fighter may be able to turn on a dime (the square corner has been solved!), become invisible to bad guys and reveal itself only to guys wearing white hats, and go warp nine forever on a teaspoon of water for fuel, but if it's available only one day out of 30 because of maintenance, then it limits your go-to-war options a bit. Logistics is no stranger to global airlifters and trash haulers, since spares and repairs constitute a lot of the business of this kind of flying. But in the test business, the reliability and maintainability considerations of full scale development (FSD) hardware and software may be something new. The maintainers and the fliers out in the operational world are the RECIPIENTS of what you pass through your tests. And by the time you get the new stuff, a lot of the reliability, maintainability, and supportability stuff has already been decided. Not all of it, but a lot of it.

OK, let's say you're finally out of Test Ops and you're in a Combined Test Force (CTF) as an Ops Officer or Director. Either way, part of your job is to see that the new Buck Rogers

equipment dreamed up by the Contractor(s) in response to the Statement of Need (SON) (remember back in the Systems Acquisition course?) fits the bill in all respects. You're the Responsible Test Organization (RTO) that's charged with the test and evaluation responsibilities. TPS has done an outstanding job of ensuring you have the tools to do your tests and make your evaluations. The defense of the American Way of Life is now in your hands. It's time to get to work.

You may very easily repeat the words of Admiral Ernest J. King, CNO, in 1942 -- "I don't know what the hell logistics is ... but I want some of it." Luckily you have the advantage of knowing what it is simply by reading the next few paragraphs.

AFR 800-8, "Integrated Logistics Support (ILS) Program," says that the objective of ILS is to "field weapon systems and equipment that achieve the required readiness and sustainability posture at an affordable life-cycle cost." Simply stated, this means that the ILS program should be able to provide weapon systems that place the bombs on target, return to base in one piece, be able to do the mission as many times as needed, and not cost the national treasury. Great, but how? Three items that are critical to these capabilities are reliability, maintainability, and supportability. Let's cover each of these "-ilities" one at a time.

Reliability means that not only does the hardware and software meet performance specs, but it's able to do the intended job often enough to make it worthwhile. That means that it'll last and perform as a system on more than one or two flights. It means that the fighter jock's fangs won't have to retract because he has to go ready another jet because the first one crumped during takeoff roll. It means that the airlifter will be able to fly anywhere anytime to deliver everything--every time. It means that the bomber can launch and leave or penetrate and launch (with a reasonable survivability probability) with the system doing it's job when required. It's a probability of the system or its component parts being ready and able to do what it's supposed to do when it or they are called upon.

Maintainability is the ability of the maintainers to inspect, troubleshoot, and repair/replace your weapon system parts when the calendar or the aircrew or the aircraft tells them it's necessary. It's accessibility (without having to remove a thousand fasteners) and it's commonality (one tool fits all).

Supportability is the ability to provide the weapon system where it's needed, when it's needed, and with all the equipment and material necessary to keep it going for as long as it's needed.

In the early days, new weapon systems being acquired by Uncle Sam were judged and bought based on their ability to meet performance specs and goals, th eir cost, and their development schedule. But, as these systems began to age through 10, 20, and 30 years, it became readily

evident that support costs, consisting of maintenance, spares, periodic overhaul, continual training, required facilities, and modifications to all the preceding, consumed more bucks than the initial price of the system itself (see Figure 1).

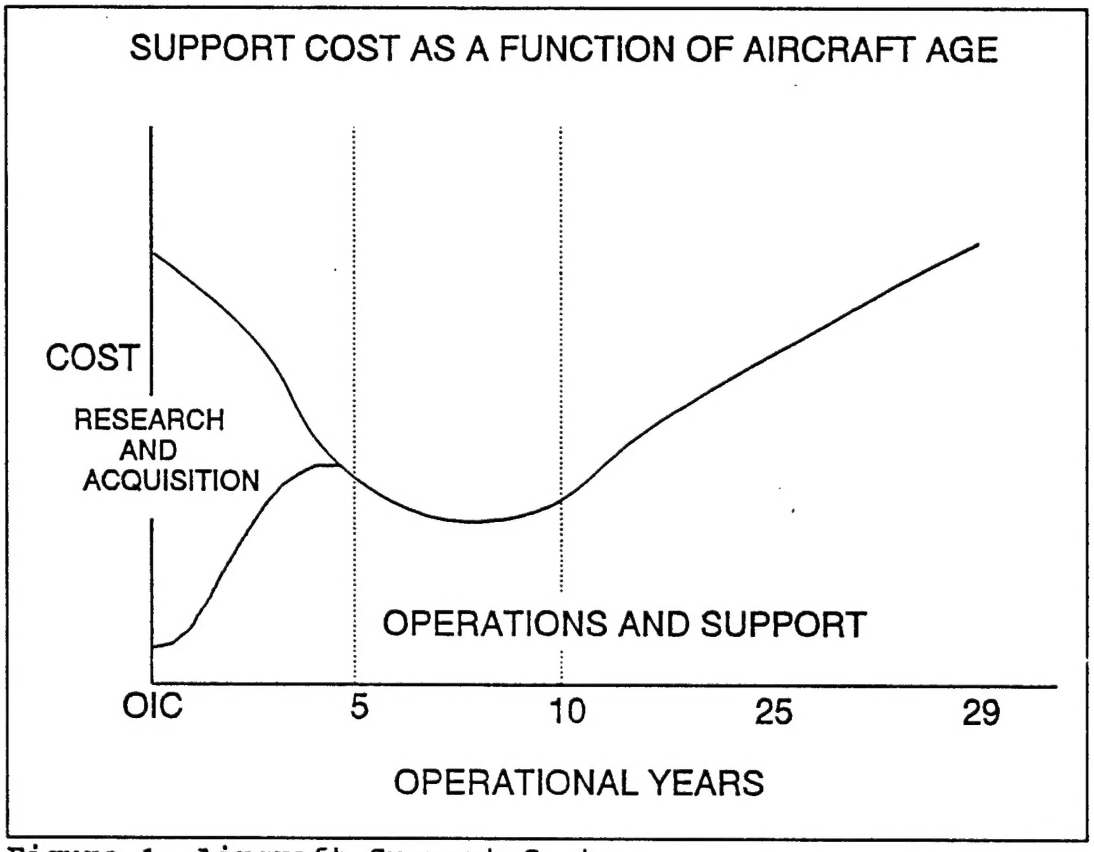


Figure 1 Aircraft Support Costs

This observation led to the development of the "life cycle cost" considerations. This cost includes everything from the initial concept exploration (back to the acquisition course again) through the demonstration/validation, through the full scale development, through the production and deployment, through the operational support, to disposal. So it includes, not only the cost of the hardware and software development and production, but also all those "incidental" items that cost many times more than development and production. What is even more interesting is that these life cycle costs were found to be determined or locked-in during the initial development stages of the program--about 70% by the end of the concept exploration phase and 90% by the beginning of the full scale development phase.

This means that the reliability, maintainability, and supportability designs and decisions have been almost entirely frozen before the guys who do the work on the systems and maintain them have even seen, used, and evaluated them. Bad designs in these operational areas will plague the users

for years. Therefore, the best way to minimize these huge costs (about 60% of the total system life cycle cost) and problems was to build the systems with the maintainers, trainers, and reliability people involved from the conceptual development phase on. Their inputs, tests, and evaluations during the early acquisition phases--before the designs and concepts are frozen--will lower the overall life cycle costs and provide a better product to the user in the field.

A good example of the old method of design for performance alone is the F-4. In 1983, the Air Force spent $250,000 and 58,000 manhours removing ejection seats from the F-4. But there was nothing wrong with the seats! Underneath them was a radio that required a lot of easy-to-do maintenance, and you had to remove the seats to get at them to do the simple, periodic maintenance. Big bucks spent downstream for something that could have been better designed and avoided at the start.

Among the examples of applying "lessons learned" are the A-10 and the ATF aircraft. Figure 2 illustrates the maintainability and accessibility built-in during the initial design of the Warthog. The ATF, as of this writing, also has "blue suit hands-on" during the demonstration/validation phase of the aircraft development so that maintainability and accessibility can be improved in the design, as a result of actual maintenance experience, prior to the building of the first full scale development aircraft.

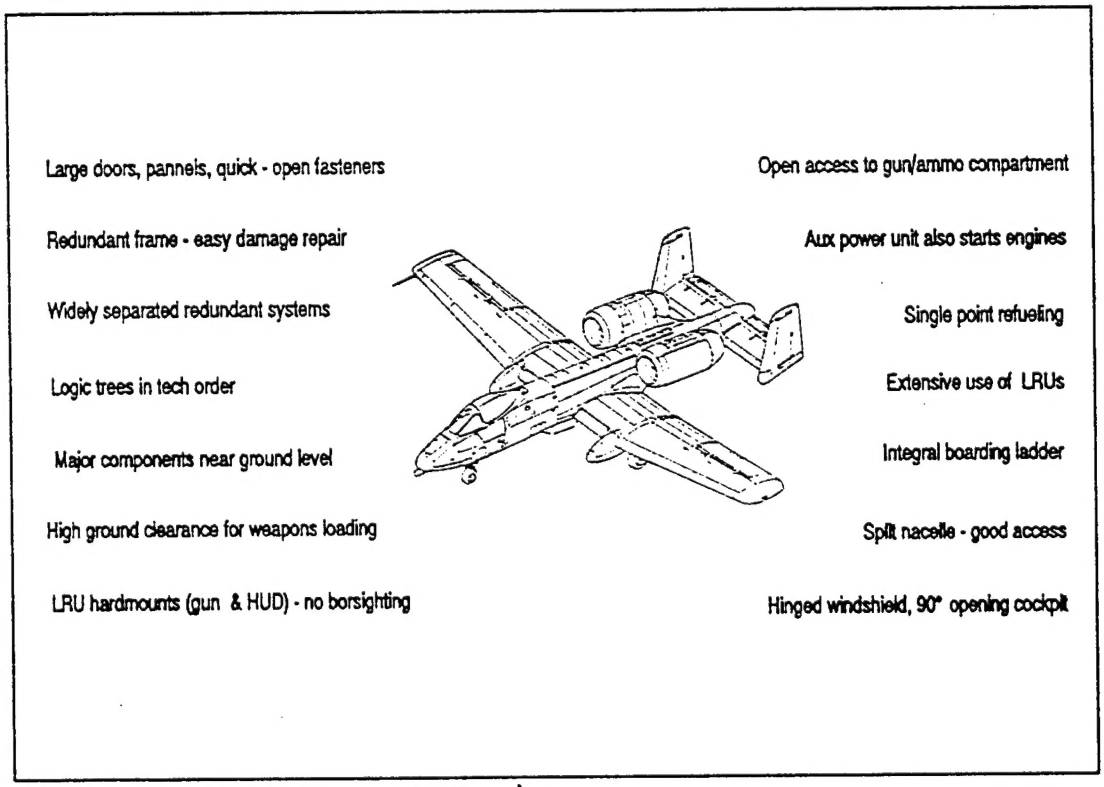


Figure 2 A-10 Support Design

Supportability is also a big consideration during design of a new weapon system. The B-2 bomber, while costing mega-bucks for each copy, was designed in part to use support equipment that has already been developed and in use for the B-52 and B-1 bombers. The munitions handling units that are used on the venerable old Buff can be used on the newest of our stealth bombers. At least the basic equipment can, with a few extra added (and removable) modifications. The savings extend to more than just the development and purchase costs of new support equipment, too. Maintenance of this support equipment takes the same training, the same number of skilled people, the same tech orders, and the same fluids, tires, hoses, etc. as those already in the field. There are no new airlift requirements for new support equipment because of this commonality and there are a lot more divert base options for the B-2 because the equipment is the same as the B-1 and the B-52. The cascade effect is the same for every piece of "off-the-shelf" hardware that can be used on new systems.

This ILS and life cycle cost involvement during the acquisition and test phases doesn't happen by accident, though. The maintainers and supportability and reliability specialists have to be brought in and involved in the system design at the start, and they have to remain a functional part of the design and test/evaluation team all the way through testing and production. The parallel activities of test and ILS are shown in Figure 3.

SYSTEM LIFE CYCLE TECHNICAL ACTIVITIES

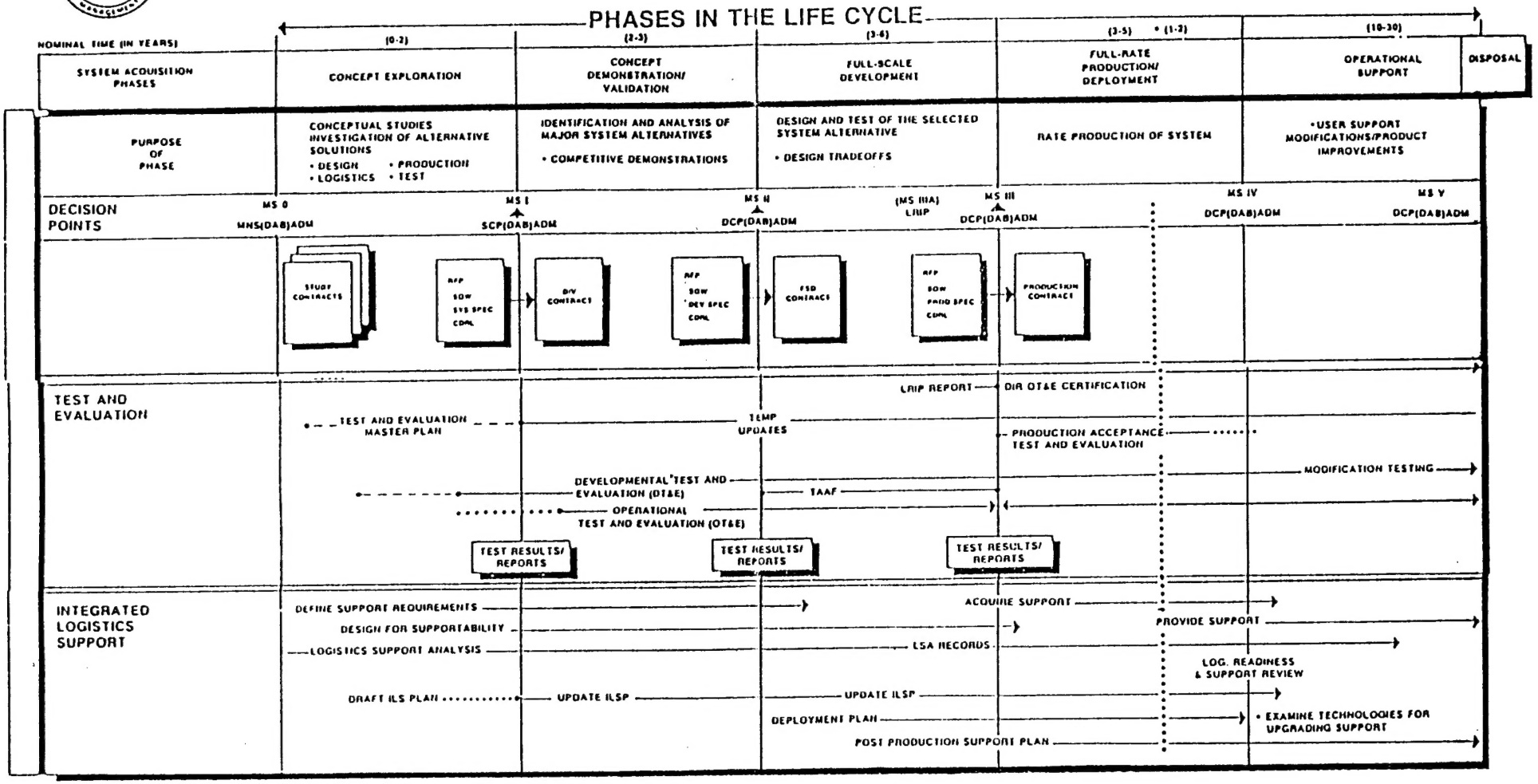


Figure 3 Parallel Life Cycle Activities

Definition of the realiability, maintainability, and supportability requirements are made by the using command in the SON. During coordination, the logistics community reviews the requirements for specificity and measurability. With the approval of the SON, Logistics Command appoints a Deputy Program Manager for Logistics (DPML) to the program. He or she works for AFLC, not AFSC. The DPML's responsibility is to ensure ILS aspects are considered and included in the weapon system design. This is accomplished through the Integrated Logistics Support Plan (ILSP) and the Logistics Support Analysis (LSA). The ILSP outlines the actions necessary to accomplish the ILS objectives of providing a reliable and supportable weapon system by describing the concepts, requirements, tasks, schedules, and subordinate plans associated with each ILS element.

The LSA is the design engineering process that actually designs supportability into the system. It includes analyses of existing systems that define high maintenance areas and the reasons for them. From these "lessons learned," the design engineers can improve readiness and supportability by increasing reliability and decreasing maintenance requirements and time. For example, reducing design stresses and providing redundancy can increase reliability, while designing quick access panels, reducing the number of fasteners, and providing visual inspection accessibility can reduce maintenance requirements.

Test and evaluation is the "proof of the pudding" for the design characteristics of the weapon system. Although much of the logistics testing takes place during OT&E, the data gathered during DT&E is very valuable. Let's briefly look at the planning guidelines and the methods of conducting logistics testing.

PLANNING GUIDELINES

1. Develop a test strategy for each logistics-related objective. Ensure that your test planning encompasses all the logistics elements. Table 1 illustrates general objectives that should be translated into detailed qualitative and quantitative requirements.

ACQUISITION PHASE / TEST TYPE	CONCEPT EXPLORATION/ DEFINITION	CONCEPT DEMONSTRATION/ VALIDATION	FULL SCALE DEVELOPMENT	FULL-RATE PRODUCTION/DEPLOYMENT	OPERATIONAL SUPPORT
DEVELOPMENT T&E	• SELECT PREFERRED SYSTEM AND SUPPORT CONCEPTS	• IDENTIFY PREFERRED TECHNICAL APPROACH, LOGISTIC RISKS, AND PREFERRED SOLUTIONS	• IDENTIFY DESIGN PROBLEMS AND SOLUTIONS IN -SURVIVABILITY -COMPATIBILITY -TRANSPORTATION -R&M -SAFETY -HUMAN FACTORS	• ASSURE PRODUCTION ITEMS MEET DESIGN REQUIREMENTS AND SPECIFICATIONS	• ASSURE ADEQUACY OF SYSTEM DESIGN CHANGES
OPERATIONAL T&E SUPPORTABILITY ASSESSMENT	• ASSESS OPERATIONAL IMPACT OF CANDIDATE TECHNICAL APPROACHES • ASSIST IN SELECTING PREFERRED SYSTEM AND SUPPORT CONCEPTS • ESTIMATE OPERATIONAL COMPATIBILITY AND SUITABILITY	• EXAMINE OPERATIONAL ASPECTS OF ALTERNATIVE TECHNICAL APPROACHES • ESTIMATE POTENTIAL OPERATIONAL SUITABILITY OF CANDIDATE SYSTEMS	• ASSESS OPERATIONAL SUITABILITY -OPERATIONAL R&M -BUILT-IN DIAGNOSTIC CAPABILITY -TRANSPORTABILITY • EVALUATE LOGISTICS SUPPORTABILITY -EFFECTIVENESS OF MAINTENANCE PLANNING -APPROPRIATE PERSONNEL SKILLS/GRADES -APPROPRIATE SPARES, REPAIR PARTS, BULK SUPPLIES -ADEQUATE SUPPORT EQUIPMENT, INCLUDING EFFECTIVE ATE AND SOFTWARE -ACCURATE AND EFFECTIVE TECHNICAL DATA; VALIDATION/ VERIFICATION OF TECHNICAL MANUALS -ADEQUATE FACILITIES (SPACE, ENVIRONMENTAL SYSTEMS, STORAGE) -EFFECTIVE PACKAGING, LIFTING DEVICES, TIE-DOWN POINTS, TRANSPORTATION INSTRUCTIONS	• ASSURE PRODUCTION ITEMS MEET OPERATIONAL SUITABILITY REQUIREMENTS	• DEMONSTRATE ATTAINMENT OF SYSTEM READINESS OBJECTIVES • UPDATE O&S COST ESTIMATES • EVALUATE OPERATIONAL SUITABILITY AND SUPPORTABILITY OF DESIGN CHANGES • IDENTIFY IMPROVEMENT REQUIRED IN SUPPORTABILITY PARAMETERS • PROVIDE DATA REQUIRED TO ADJUST ILS ELEMENTS

*7-2-MCL2-000186-0

Table 1 ILS OBJECTIVES IN THE T&E PROGRAM

SOURCE: DSMC, INTEGRATED LOGISTICS SUPPORT GUIDE, MAY 1986

2. Incorporate the logistics test requirements into the formal DT&E/OT&E plans.

3. Identify tests for logistics elements that will be performed outside of the normal DT&E and OT&E. These may include subsystems that require off-system evaluation.

4. Identify required resources, including the number of test articles, support items, training, and skilled personnel, to support testing.

5. Establish an operationally realistic test environment, both for the equipment under test and for the people performing the testing. This environment includes tech manuals and procedures for performing maintenance and personnel representative of those who will eventually operate and maintain the fielded system.

6. Ensure the operational testing will provide sufficient data on high cost and high maintenance burden items. For example, early tests for high cost critical spares can yield resuits used to reevaluate selections.

7. Ensure the Test and Evaluation Master Plan (TEMP) includes logistics testing.

8. Identify the planned use of logistics data during the evaluation to avoid mismatch of data collection and analysis requirements.

CONDUCTING LOGISTICS TESTING

Remember that the purposes of conducting the logistics tests are to

- measure the supportability of a developing system throughout the acquisition process
- identify supportability deficiencies and possible corrections or improvements as the data is analyzed
- assess the operational effectiveness of the planned support system.

The contractor's maintenance personnel usually have the majority of the work at the start of DT&E, since the new system normally has only contractor procedures for the maintenance and repair of the system. The hardware and the software are far enough along in development to allow testing, but the "start-up bugs" are still being worked out. Few tech orders are usually available for blue-suit maintainers, and those that are have yet to be validated or verified (more on this later). The Air Force maintenance crews are part of the test team, but are usually there

to look over the shoulders of the contractor maintainers, to observe, and to learn. Data generated from this part of the testing are still important.

There are several specific logistics tasks that are included during the conduct of a system's tests and evaluations for technical and operational effectiveness and suitability.

First, the system performance is being tested to see if it works at all and, if it does work, how well it does what it's supposed to do. Test objectives and procedures are written to accomplish this performance evaluation, on the ground and in the air. Analysis is done on the data derived from instrumentation parameters that are recorded during testing and the objectives are passed or failed. In addition to the performance data, reliability data are also being collected. For example, the environment in which the electronics, ordnance, controllers, actuators, and computers are located can (and should be) measured. These data allow the predicted vibration, acoustic, temperature, and other environmental conditions to be validated or updated. Future failures of this equipment will require these data for reliability and redesign information. But the instrumentation for these data have to be specified, designed, installed, and maintained during the test program. That's part of your job as a tester.

Second, failure data of the full scale development equipment is recorded and tracked. This is sometimes a point of contention between the contractor and the Air Force logistics specialists. The contractor points out that the equipment under test is not in production configuration, since it is still full scale development, and the failure tracking at this point is comparing "apples and oranges." The Air Force maintains that the failures of the equipment occur as a result of design deficiencies or random failures, resulting in data that can be used to improve the design of the full scale development equipment prior to production configuration. The result is that the data is normally tracked, but its applicability to production configured hardware and software is limited.

Tracking of logistics data is normally done via the Systems Effectiveness Data System (SEDS), which is a computerized system into which failure data are coded and entered. These data are updated for cause of the failure, the repair time and the corrective action as the FSD testing continues. The data are also used to start the determination of the reliability of the system equipment in order to determine spares requirements and warranty provisions of the production contract. The contractor and the blue suit maintainers both collect SEDS data.

While at first glance, this tracking and "bean counting" may appear rather simple, particularly in this age of computers, consider the future outlook for aircraft avionics. In the upcoming fighter series, the black boxes that control the flight performance, weapons delivery, navigation, controls and displays, and communications won't be the black boxes you're used to seeing today.

They may be a collection of modules, each identical to the other, that have multiple redundancy for all of the functions listed above. In other words, when one module fails in the performance of one function, the central processor(s) will detect the failure, shift that function to another module and reassign another function, one that it still can perform, to the original module. It doesn't take a rocket scientist to see that, even though the onboard computers record the failure (and perhaps even the reason for it), the reliability tracking of modules will be radically altered. Did the first unit really "fail" since it was able to be reassigned another function and it continued to operate? Does it have to even be replaced, since the multiple redundancy in identical modules still allows many functional backups?

You, as an aircraft driver or test engineer, really only see the increased reliability of the weapon system to perform its job. You - the test director - must evaluate the reliability and the test methods of these new systems.

Other specific logistics-related tasks during testing include:

- analysis of test results to verify achievement of supportability requirements.

- determination of improvements in supportability and related design parameters needed to meet thresholds and goals.

- identification of areas where thresholds and goals have not been demonstrated within acceptable confidence levels. (remember your course in probability and statistics?)

- development of corrections for supportability problems. These may include modifications to hardware, software, support plans, spares, or even operational tactics.

- projection of changes in costs, readiness, and logistic support resources due to the corrections you just made above.

Finally, there is a beast called the Joint Reliability/Maintainability Evaluation Team, or JRMET, which meets regularly during and after the test and evaluation phase of the acquisition process. This team is comprised of the System Program Office (SPO), the logistics support members of your test team, and the contractor. On a periodic basis (usually quarterly), the JRMET meets to:

- evaluate and summarize the logistics data collected and entered via the SEDS.

- identify supportability and reliability problem areas in the developing system.

- air differences in data collection and evaluation methods and conclusions.

- propose possible corrections and/or modifications to the identified system problem areas.

- initiate the system's data base for reliability and supportability throughout its life cycle.

What the JRMET does not do is evaluate whether the system meets the technical requirements imposed by the specification outside of the logistics area. It is a useful tool for the user to start gaining an insight into the possible logistics-related problem areas and the solutions before the system is fielded for operational use.

One more item to mention in this overview--technical orders, or manuals for the proper use and maintenance of the system equipment. These books range from your flight manuals to the periodic maintenance and repair of each piece of equipment in the system, with diagrams, charts, figures, and tables for everything from drag counts to diagnostic troubleshooting of problems. The tech orders are usually prepared by the contractor as part of the basic contract and the aircrews, trainers, and maintainers all review, critique, and approve the documents before they're released to the field. The review usually takes place in three steps:

1. Tabletop review--this is a meeting away from the equipment in which the manuals are reviewed for basic understandability and technical content coverage and accuracy.

2. Tech order validation--this is a hands-on, step-by-step accomplish-ment of the procedures by the contractor personnel (usually, but not always, with the blue-suit folks looking over their shoulders). This is the opportunity for the contractor wrench-benders and engineers to "validate" the steps and the results of the procedures before passing them on to the Air Force.

3. Tech order verification--this is the Air Force's turn to perform the procedures step by step to ensure that the methods described are accurate, in the correct order, are understandable, accomplish what they're supposed to, and can be performed with the forecast skilled personnel.

Tech order validation and verification (called "TOV&V") must be done during developmental testing now, since a lot of the production decisions are being made before the conclusion of testing. Obviously, the folks in the field have got to have the repair and maintenance manuals at least when the equipment arrives on their base. The TOV&V precedes the manual publication and release, and this puts it right into your developmental tests. This causes a few problems, however, since the pubs are being reviewed and updated for final release based on the instrumentation-covered, not-necessarily-production-configured equipment. Basically, everything is having to happen at the same time, and this sometimes means that the tech orders are released for systems that are somewhat different in configuration than what the field units receive. Manual updates for production configuration are almost always necessary in today's contracting world.

While there are volumes written about logistics testing and methods of evaluating all the reliability, maintainability, and supportability factors, this brief synopsis serves only to introduce the TPS student to some of the considerations useful in setting up and conducting these types of tests. The appendices contain more information about the calculations and terminology used in determining system operational effectiveness and suitability. They're for your reference down the road when you're establishing your test program for the newest piece of cosmic hardware or software. They cover more details about reliability, maintainability, availability, statistical concepts, reliability test planning, and data base considerations, and will serve as a good refresher for you when you need it later on.

APPENDIX A

RELIABILITY

Reliability is a term used to describe quantitatively how failure-free a system is likely to be during a given period of operation. The ability to express reliability numerically is crucial, because it enables us to concretely identify the user's needs, contractual specifications, test guidelines and performance assessment.

DEFINITION OF TERMS AND CONCEPTS

Reliability

Reliability is defined as the probability that an item will perform its intended function for a specified interval under stated conditions. This definition does not specifically consider the effect of the age of the system.

The following adaptation is useful for systems that are repairable. Reliability, for repairable systems, is the probability that an item will perform its intended function for a specified interval, under stated conditions, at a given age, if both corrective and preventive maintenance are performed in a specified manner.

If a system is capable of performing multiple missions, or if it can perform one or more of its missions while operating in a degraded condition or if the mission test profiles represent only typical usage, then, the concept of a unique mission reliability becomes difficult to define. In such cases, it is preferable to use a reliability measure that is not based solely on the length of a specified time interval but rather on the definition of a specific mission profile or set of profiles.

The meaning of the terms "stated conditions" and "specified interval" are important to the understanding of reliability. The term "stated conditions" refers to the complete definition of the scenario in which the system will operate. For a ground combat vehicle, these conditions include climatic conditions, road surface, and loads that would be experienced during a selected mission profile. These conditions should reflect operational usage. The term "specified interval" refers to the length of the mission described in a mission profile. this interval may include multiple factors. For example, an air defense system mission profile will define an interval containing X rounds fired, Y hours of electronics on-time and z miles of travel. For a simpler system, say an air-burst artillery round, the interval may include a single event--round detonation.

Mean Time Between Failures

Mean time between failures (MTBF) is defined as the total functioning life of a population of an item during a specific measurement interval, divided by the total number of failures within the population during that interval. MTBF can be interpreted as the expected length of time a system will be operational between failures. The definition is true for time, cycles, miles, events, or other measure-of-life units. These various measure-of-life units permit the MTBF term to be tailored to the reliability requirements of a specific system. Some examples of this tailoring are:

- Mean rounds between failure (MRBF)
- Mean miles between operational mission failure (MMBOMF)
- Mean time between unscheduled maintenance actions (MTBUMA)
- Mean rounds between any maintenance actions (MRBAMA)

Failure Rate

Failure rate is defined as the number of failures of an item per measure-of-life unit (e.g., cycles, time, miles or events as applicable). This measure is more difficult to visualize from an operational standpoint than the MTBF measure, but is a useful mathematical term which frequently appears in many engineering and statistical calculations. As we will se in later chapters the failure rate is the reciprocal of the MTBF measure, or

$$\text{Failure Rate} = \frac{1}{\text{MTBF}}$$

SYSTEM RELIABILITY DESIGN OBJECTIVES

There are two very different system reliability design objectives. One is to enhance system effectiveness; the other is to minimize the burden of owning and operating the system. The first objective is addressed by means of mission reliability, the second by means of logistics--related reliability. Measures of mission reliability address only those incidents that affect mission accomplishment. Measures of logistics-related reliability address all incidents that require a response from the logistics system.

Mission Reliability

Mission reliability is the probability that a system will perform mission essential functions for a period of time under the conditions stated in the mission profile. Mission reliability for a single shot type of system, i.e., a pyrotechnic device, would not include a time period constraint. A system with a high mission reliability has a high probability of successfully completing the defined mission.

Measures of mission reliability address only those incidents that affect mission accomplishment. A mission reliability analysis must, therefore, include the definition of mission essential functions. For example, the mission essential functions for a tank might be to move, shoot and communicate. More specific requirements could specify minimum speed, shooting accuracy and communication range.

Logistics (Maintenance/Supply) Related Reliability

Logistics related reliability measures, as indicated above, must be selected so that they account for or address all incidents that require a response from the logistics system.

Logistics related reliability may be further subdivided into maintenance related reliability and supply related reliability. These parameters respectively represent the probability that no corrective maintenance or the probability that no unscheduled supply demand will occur following the completion of a specific mission profile.

The mathematical models used to evaluate mission and logistics reliability for the same system may be entirely different.

RELIABILITY INCIDENT CLASSIFICATION

An understanding of the relationships existing between the above reliability measures and other terms is essential to the knowledgeable application of these parameters. Figure A-1 illustrates the effects of these relationships not their causes. For example, system failures may be caused by the hardware itself, by the operator, or by inadequate/faulty maintenance.

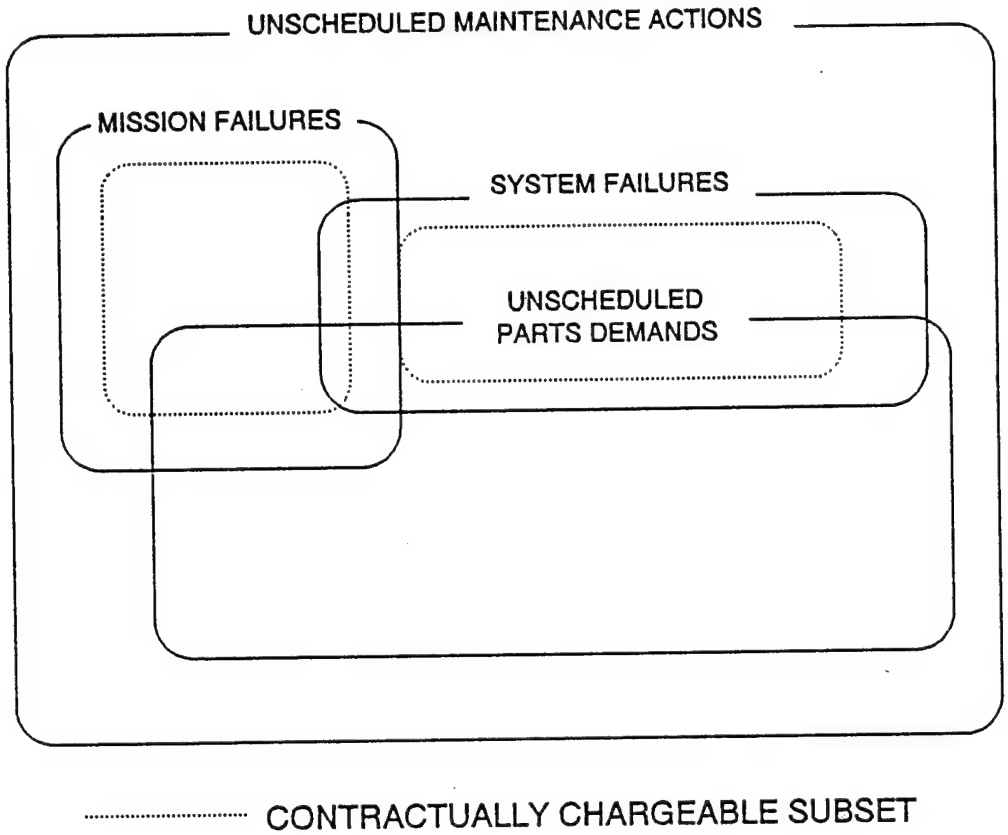


FIGURE A-1 RELIABILITY INCIDENT CLASSIFICATION

Mission Failures

Mission Failures are the loss of any of the mission's essential functions. Along with system hardware failures, operator errors and errors in publications that cause such a loss are included in this region. Mission failures are related to mission reliability measures because they prevent complete mission accomplishment.

System Failures

System failures are hardware malfunctions: they may or may not affect the mission's essential functions, and they may or may not require spares for correction. A system failure generally requires unscheduled maintenance so system failures heavily influence maintenance-related reliability.

Unscheduled Spares Demands

Unscheduled spares demands are used to evaluate supply-related reliability. All unscheduled

spares demands require a response from the supply system, so they form the basis for evaluating supply-related reliability.

System/Mission Failures Requiring Spares

System/mission failures that require spares for correction are the most critical. Mission, maintenance and supply reliabilities are affected, and the system runs the risk of being held in a non-mission-ready status for an extended period of time by logistics delays.

Contractually Chargeable Failures

Contract requirements are often established for the subset of mission failures and/or system failures for which the contractor can be held accountable. Normally excluded from contractual chargeability are such failure categories as: operator or maintenance errors; item abuse; secondary failures caused by another (primary) failure; and failures for which a "fix" has been identified (but not incorporated in the test article that failed). It should be noted tat, in operation, all failures (in fact, all unscheduled maintenance actions) are relevant regardless of contractual chargeability, and should therefore by included in operational evaluations.

SYSTEM RELIABILITY MODELS

System reliability models are utilized to describe visually and mathematically the relationship between system components and their effect on the resulting system reliability. A reliability block diagram or structural model provides the visual representation while the mathematical or "math" model provides the analytical tool to calculate quantitative reliability values.

The following notation is used in the discussion of reliability models:

R_s = reliability of the system

R_i = reliability of the i^{th} subsystem

Q_s = $1 - R_s$ = unreliability of the system

Q_i = $1 - R_i$ = unreliability of the i^{th} subsystem

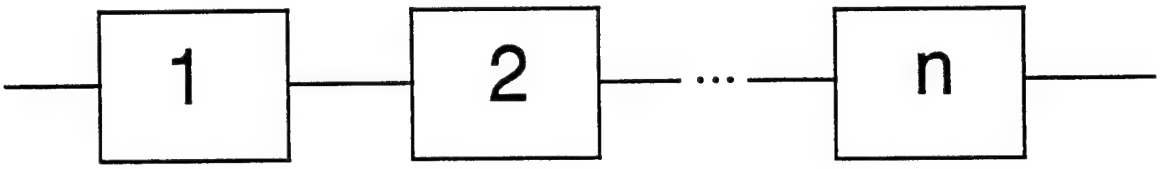
Π = product of (Note: This operator is used in the same fashion as Σ for summation, but it indicates multiplication rather than addition).

NOTE: In the following discussion it is assumed that all subsystems function independently of one another, that is, failures of different subsystems are statistically independent of each other. For many systems this represents a realistic assumption. the reliability analysis for dependent subsystems is significantly more complex. Independent operation, practically speaking, means that a failure of one system will not cause a change in the failure characteristics of one or more

other subsystems. Therefore, replacement of the single failed subsystem should permit continued operation of the entire system, because other subsystems were not affected.

Series Model

When a group of components or subsystems is such that all must function properly for the system to succeed, they are said to be in series. A system consisting of a series arrangement of n subsystems is illustrated in the following block diagram:



The mathematical model is

$$R_s = R_1 R_2 \ldots = \prod_{i=1}^{n} R_i$$

Redundant Models

The mission reliability of a system containing independent subsystems can usually be increased by using subsystems in a redundant fashion, that is, providing more subsystems than are absolutely necessary for satisfactory performance. The incorporation of redundancy into a system design and the subsequent analysis and assessment of that design is a complex task and will not be addressed here in detail. We will define the elements of redundancy and present several simplified examples.

Redundancy Characteristics. Redundancy can be defined by three basic characteristics.

- First, the level at which redundancy is applied. For example, we could have redundant piece parts, redundant black boxes, or complete redundant systems.

- Second, the operating state of the redundant element. The redundant part, subsystem, etc., may exist in the circuit as an active functioning element or as a passive, power off, element. for example, an airport that maintains two separate operating ground control approach radars at all times has active redundancy for that capability. Carrying a spare tire in your trunk is an example of passive redundancy.

- Third, the method used to activate the redundant element. Consider the passive redundancy case of the spare tire. The vehicle driver represents the switching device that decides to activate the spare. Obviously mission time is lost in installing the spare. the opposite case is represented by the use of an electronic switching network that senses the failure of Box A and automatically switches to Box B without lost time or mission interruption.

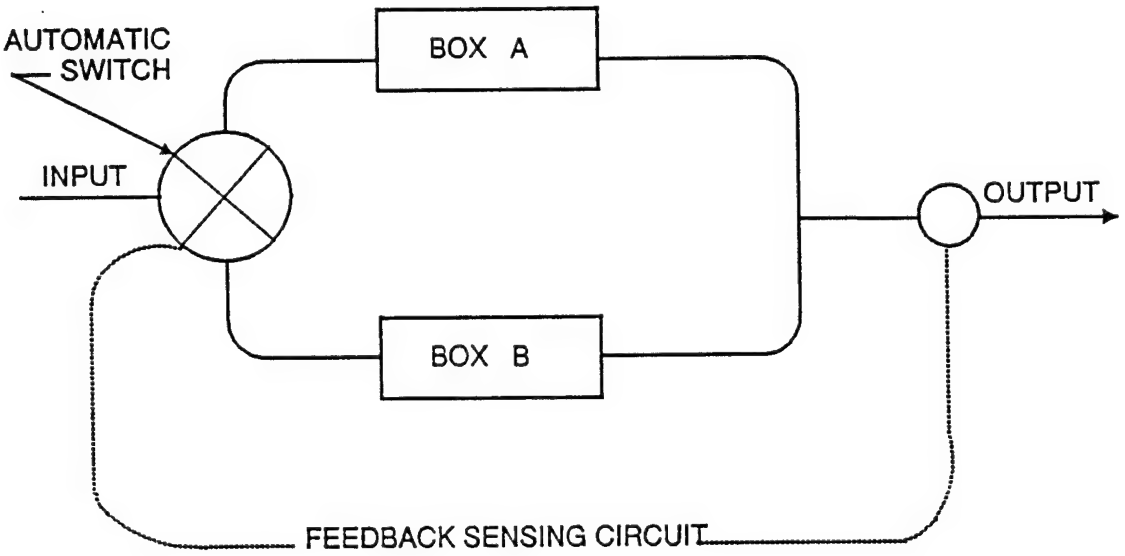
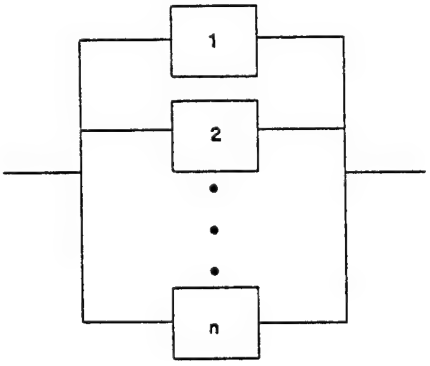


FIGURE A-2 PASSIVE REDUNDANCY WITH AUTOMATIC SWITCHING

Our examples will consider only simple active redundancy. In this type o redundancy, all the operable subsystems are operating, but only one is needed for satisfactory performance. There are no standby subsystems, and no repair is permitted during the mission. Such a system of n subsystems is illustrated in block diagram form as:



Note: Simple active redundancy requires that only one of the n subsystems be operating for misssion success.

The mathematical Model is

$$Q_s = Q_1 Q_2 \ldots Q_n = \prod_{i=1}^{n} Q_i = \prod_{i=1}^{n} (1 - R_i)$$

$$R_s = 1 - Q_s = 1 - \prod_{i=1}^{n} (1 - R_i)$$

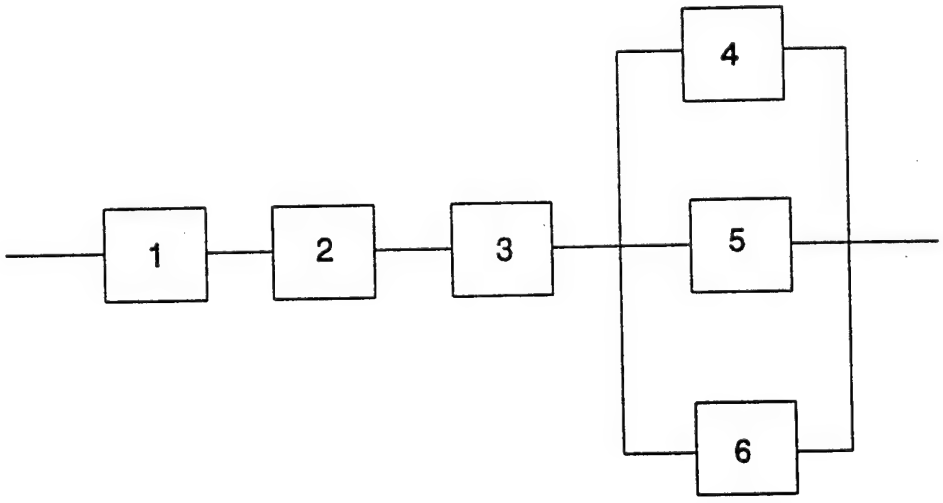
This model again assumes that there is statistical independence among failures of the subsystems. This assumption is important because dependence between subsystems can have a significant effect on system reliability. Calculations based on an assumption of independence can be erroneous and misleading. In fact, erroneously assuming failure independence will often result in overestimating system reliability for an active redundant system and underestimating reliability for a series system.

Implications of Redundant Design. While redundant design does improve mission reliability, its use must be weighed against the inherent disadvantage. These disadvantages include greater initial cost, increased system size and weight, increased maintenance burdens and higher spares demand rates. These factors must be considered by using and procuring agencies and by testing organizations when assessing the true mission capability and support requirements.

Although there are some possible exceptions, redundancy generally improves mission reliability and degrades logistics reliability. Case Study 2-3 gives a numerical comparison between mission- and maintenance-related reliability.

Mixed Models

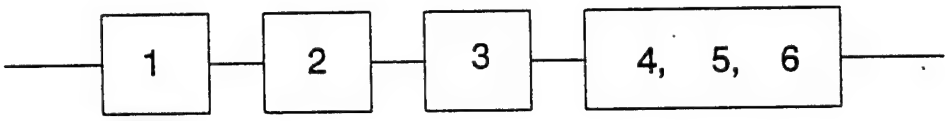
One system configuration that is often encountered is one in which subsystems are in series, but redundancy (active) is applied to a certain critical subsystem(s). A typical block diagram follows:



This type of model (or any mixed model, for that matter) is characterized by working from low to high levels of assembly. In this case, assuming independence and active redundancy, we can apply the following equation:

$$R_{4,5,6} = 1-(1-R_4)(1-R_5)(1-R_6)$$

We can now represent the redundant configuration of 4, 5, and 6 by a single block on the diagram.

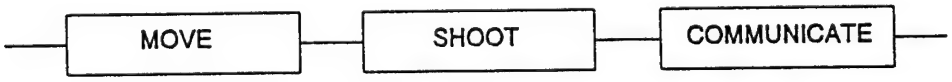


$$R_s = R_1R_2R_3R_{4,5,6}$$

$$R_s = R_1R_2R_3[1-(1-R_4)(1-R_5)(1-R_6)]$$

Functional Models

The series, redundant and mixed models mentioned above, are hardware-oriented in that they display hardware capabilities. In some cases, it is desirable to model a system from a functional standpoint. As an example, the functional reliability block diagram for a tank is shown below:



The functions may be defined as:

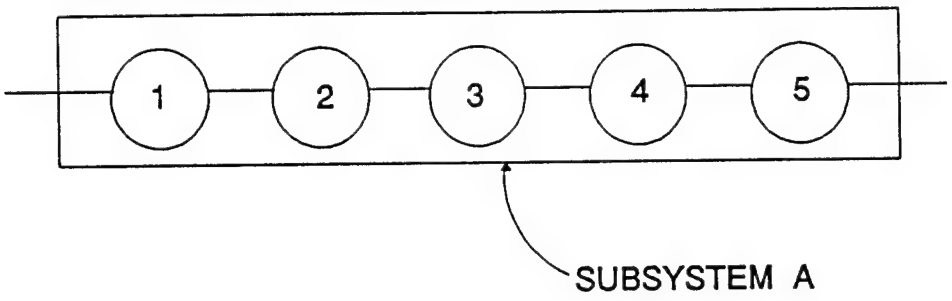
- MOVE. The mobility system must be capable of effectively maneuvering such that the system can maintain its assigned position within a tactical scenario. Specific requirements are determined for speed, turning, terrain, etc.
- SHOOT. The main gun must be capable of delivering effective fire at the rate of X rounds per minute.
- COMMUNICATE. The communication system must be capable of providing two-way communication with other vehicles and with fixed stations within specific ranges and terrain confines.

Note that this concept addresses mission-essential functions, but in no way implies how these functions will be accomplished. Generally the functional model approach is helpful in the program formulation stages of a program when specific hardware information is not necessary and frequently not desired. This type of model can provide a useful transition from operational requirement to engineering specifications.

RELIABILITY ALLOCATION

The previous section presented the topic of functional reliability models and indicated that these models provided a useful means of transitioning from operational requirements to engineering specifications. the process of transitioning from operational requirements to engineering specifications is known as reliability allocation. The reliability allocation process "allocates" the reliability "budget" for a given system or subsystem to the individual components of that system or subsystem. An example will prove helpful.

Suppose we have previously determined that the reliability of an electronic subsystem A, must equal or exceed 0.90, and that this subsystem has been designed with 5 parts all functioning in series. For this example, we will assume Parts 1, 2 and 3 are the same and the best available piece part for Part 4 has a reliability of 0.990. How can we allocate the reliability budget for this subsystem to its individual parts?



Using the previous equation we have

$$R_{Total} = R_1R_2R_3R_4R_5$$

$$= R_1R_2R_3(0.990)R_5$$

Solving for $R_1R_2R_3R_5$ we have

$$R_1R_2R_3 = \frac{0.900}{0.990} = 0.909$$

If we assume $R_1 = R_2 = R_3 = R_5$ then,

$$R_1 = R_2 = R_3 = \sqrt[4]{0.909} = 0.976$$

If we can locate piece parts for Part 5 with $R_5 = 0.985$, then

$$R_1R_2R_3 = \frac{0.909}{0.985} = 0.923. \quad \text{So}$$

$$R_1 = R_2 = R_3 = \sqrt[3]{0.923} = 0.923$$

Summary of Allocation

Case I	Case II
$R_1 = R_2 = R_3 = R_5 = 0.976$	$R_1 = R_2 = R_3 = 0.973$
$R_4 = 0.990$	$R_4 = 0.990$
	$R_5 = 0.985$

Another, and somewhat more frequently used approach to reliability allocation is one in which reliability is allocated on the basis of allowable failures or failure rates.

The understanding of reliability allocation is important to those individuals who must be concerned with hardware operating characteristics below the system level. This is especially true to development and testing organizations who are frequently faced with predicting system performance early in development, when no full-up system exists but when subsystem or component test data may be available.

APPENDIX B
MAINTAINABILITY

INTRODUCTION

Maintainability and reliability are the two major system characteristics that combine to form the commonly used effectiveness index--availability. While maintainability is important as a factor of availability, it also merits substantial consideration as an individual system characteristic. the importance of this parameter in the national defense posture becomes even more obvious when we consider that at least one branch of the armed services spends one-third of its budget on system maintenance activities.

Several aspects of system maintainability must be addressed before an accurate assessment can be undertaken. First, the difference between maintainability and maintenance must be understood. Maintainability is a design consideration, whereas maintenance is the consequence of design. The maintenance activity must live with whatever maintainability is inherent in the design, that is, it must preserve the existing level of maintainability and can do nothing to improve that level. Maintenance is therefore defined as "all actions necessary for retaining a hardware item in or restoring it to an optimal design condition." The second consideration is that maintainability requirements can be specified, measured and demonstrated. Unlike reliability, detailed and quantitative study of maintainability was not initiated until the early 1950s. Until recently, maintainability often was viewed as a "common sense" ingredient of design. It is now seen as a factor of the design process and an inherent design characteristic that is quantitative in nature and therefore lends itself to specification, demonstration, and trade-off analysis with such characteristics as reliability and logistics support.

DEFINITION OF TERMS AND CONCEPTS

Maintainability

Maintainability is defined as a characteristic of design and installation. This characteristic is expressed as the probability that an item will be retained in, or restored to, a specified condition within a given period if prescribed procedures and resources are used.

A commonly used working definition states that Maintainability is a design consideration. It is the inherent characteristic of a finished design that determines the type and amount of maintenance required to retain that design in, or restore it to, a specified condition.

Maintenance

This term is defined as all actions required to retain an item in, or restore it to, a specified condition. This includes diagnosis, repair inspection.

Preventive Maintenance

This term is defined as systematic inspection, detection and correction of incipient failures either before they occur or before they develop into major defects. Adjustment, lubrication and scheduled checks are included in the definition of preventive maintenance.

Corrective Maintenance

This term is defined as that maintenance performed on a non-scheduled basis to restore equipment to satisfactory condition by correcting a malfunction.

CONSIDERATIONS IN PLANNING MAINTAINABILITY ASSESSMENT

An understanding of the principal elements of maintainability is essential to the assessment planning process. Certain elements of a design basically define a system's inherent maintainability and thus determine the related maintenance burden and affect system availability.

It is apparent, from the definition of maintainability, that the ability and need to perform maintenance actions is the underlying consideration when assessing maintainability. The factors which affect the frequency with which maintenance is needed are reliability and the preventive maintenance schedule. Those which affect the ability to perform maintenance on a given weapon system may be broken down into three categories: the physical design of the system, the technical personnel performing the maintenance and the support facilities required.

The consideration of maintenance when designing a system is not new. There have been very successful efforts in the development of automatic check out and design for accessibility, etc. What is new is the emphasis on quantitative treatment and assessment which results in a complete change in design philosophy, design approach, and design management. In the past, design for "maximum" or "optimum" reliability and maintainability was emphasized. This resulted in "unknown" reliability and maintainability. New techniques permit us to bring qualitative design judgments into an area of quantitative measurement. We can thus establish quantitative design goals and orient the design to specific mission thresholds, not to "optimum" or "maximum" goals. Maintainability design considerations and testing intended to assess system maintainability characteristics must be based on established quantitative requirements (thresholds and goals). In addition to verifying these values, the maintainability test and evaluation program also should address the impact of physical design features and maintenance action frequency on system maintenance.

Some physical design features affect the speed and ease with which maintenance can be performed. These features and pertinent questions are:

- Accessibility: Can the item to be repaired or adjusted be reached easily?

- Visibility: Can the item being worked on be seen?
- Testability: Can system faults be detected and isolated to the faulty replaceable assembly level?
- Complexity: How many subsystems are in the system? How many parts are used? Are the parts standard or special-purpose?
- Interchangeability: Can the failed or malfunctioning unit be "swapped around" or readily replaced with an identical unit with no need for recalibration?

In addition to the listed physical design factors, the frequency with which each maintenance action must be performed is a major factor in both corrective and scheduled or preventive maintenance. Thus, reliability could have a significant impact on corrective maintenance, and such design features as self-check-out, reduced lubrication requirements, and self-adjustment would affect the need for preventive maintenance.

Personnel and human factor considerations are of prime importance. These considerations include the experience of the technician, training, skill level, supervision required, supervision available, techniques used, physical coordination and strength, and number of technicians and team work requirements.

Support considerations cover the logistics septum and maintenance organization required to support the weapon system. They include availability of supplies, spare parts, technical data (TOs and manuals), built-in test equipment, external test equipment and required tools (standard or special) and servicing equipment.

While some elements of maintainability can be assessed individually, it should be obvious that a true assessment of system maintainability generally must be developed at the system level under operating conditions and using production configuration hardware.

QUANTITATIVE MAINTAINABILITY INDICES

The following paragraphs describe the various mathematical indices used to quantify maintainability. It is important to remember, however, that these relationships merely categorize data derived from planned testing. For maintainability, the test planning phase is equal in importance to the assessment phase. Testing that does not adequately demonstrate the effect of the above physical design features and personnel and support aspects provides data that effectively conceal the impact of these critical elements.

Indices used to support maintainability analysis must be composed of measurable quantities, must provide effectiveness-oriented data and must be readily obtainable from operational and applicable

development testing. If they are, system designers, users and testers can evaluate candidate system characteristics and logistics and maintenance practices more precisely.

Mean-Time-to-Repair (MTTR) or Mct

MTTR is the total corrective maintenance down time accumulated during a specific period divided by the total number of corrective maintenance actions completed during the same period. MTTR commonly is used as an on-equipment measure but can be applied to each maintenance level individually. The MTTR considers active corrective maintenance time only. Because the frequency of corrective maintenance actions and the number of man-hours expended are not considered (clock hours are used), this index does not provide a good measure of the maintenance burden.

Maximum-Time-to-Repair (MaxTTR) or MmaxC

MmaxC is the maximum corrective maintenance down time within which either 90 or 95 percent (as specified) of all corrective maintenance actions can be accomplished. An Mmax C requirement is useful in those special cases in which there is a tolerable down time for the system. Ideally, we would like to be able to state an absolute maximum, but this is impractical because there will inevitably be failures that require exceptionally long repair times. A 95th percentile MmaxC specification requires that no more than 5 percent of all corrective maintenance actions take longer than MmaxC.

Maintenance Ratio (MR)

MR is the cumulative number of man-hours of maintenance expended in direct labor during a given period of time, divided by the cumulative number of end-item operating hours (or rounds or miles) during the same time. The MR is expressed at each level of maintenance and summarized for all levels of maintenance combined. Both corrective and preventive maintenance are included. Man-hours for off-system repair of replaced components and man-hours for daily operational checks are included for some classes of systems.

Particular care must be taken that the operating hour base be clearly defined. for example, in the case of combat vehicles, either system operating hours or engine hours could be used.

The MR is a useful measure of the relative maintenance burden associated with a system. It provides a means of comparing systems and is useful in determining the compatibility of a system with the size of the maintenance organization.

For fielded systems, the MR is useful in maintenance scheduling. Some care must be exercised in relating the MR to maintenance costs, because an in-house maintenance organization will have a fixed labor cost, independent of the amount of actual use of the system, but principally fixed by the size of the maintenance staff.

Mean-Time-Between-Maintenance-Actions (MTBMA)

MTBMA is the mean of the distribution of the time intervals between either corrective maintenance actions, preventive maintenance actions or all maintenance actions. This index is frequently used in availability calculations and in statistically-oriented maintenance analyses.

Average Number of Technicians Required

The average number of technicians required at each maintenance level provides a quantitative means of expressing the personnel aspects of the overall maintenance concept. This index also provides a conversion factor from active down time to labor hours.

Off-System Maintainability Indices

The indices MTTR, MmaxC and MR all specifically exclude off-system maintenance actions. Off-system measures are particularly important if a system's maintenance concept involves extensive use of modular removable and replacement, since this type of concept transfers the maintenance burden to off-systems maintenance. As an assessment tool, off-system maintainability measures are essential. Without them, it is not possible to assess the ability of combat environment of off-system repair and logistics capability to maintain the system. Because of the system-peculiar nature of these parameters, none are specified here. Suffice it to say that a complete set of on- and off-system indices is required to adequately assess system maintainability and total maintenance burden.

Annual Support Cost (ASC)

This is the direct, annual cost of maintenance personnel, repair, parts and transportation for all corrective (either on-system, off-system or both) and preventive maintenance actions when the system operates X hours per year during the Nth year of M years service life, where the system is defined as Y units of item A, Z units of item B, etc.

The ASC provides another means of quantifying the maintenance burden of a system. The unique feature of the ASC measure is the recognition that maintenance requirements may not be uniform over the life of a system. For example, a combat vehicle will experience a high-cost year when its engine requires replacement or overhaul. this measure provides a means of interrelating durability requirements and policies for scheduled maintenance.

DIAGNOSTIC SYSTEMS

Introduction

One aspect of maintainability that has received significant attention in recent system designs is the use of automatic diagnostic systems. These systems include both internal or integrated diagnostic systems, referred to as built-in-test (BIT) or built-in-test-equipment (BITE), and external diagnostic systems, referred to as automatic test equipment (ATE), test sets or off-line

test equipment. The following discussion will focus on BIT but most of the key point apply equally to other diagnostic systems.

Need for Automatic Diagnostic Systems - BIT

As technology advances continue to increase the capability and complexity of modern weapon systems, we are relying more on the use of automatic diagnostics, i.e., BIT, as a means of attaining the required level of failure detection capability. Our need for BIT is driven by operational availability requirements which do not permit the lengthy MTTRs associated with detecting and isolating failure modes in microcircuit technology equipment. We also find that because BIT operates within the basic system and at the same functioning speed, it therefore affords us the capability to detect and isolate failures which conventional test equipment and techniques could not provide. Finally, a well designed BIT system can substantially reduce the need for highly trained field level maintenance personnel by permitting less skilled personnel to locate failures and channel suspect hardware to centralized intermediate repair facilities which are equipped to diagnose and/or repair defective hardware.

As we shall discuss, Bit is not a comprehensive solution to all system maintenance problems but rather a necessary tool required to deal with the complexity of modern electronic systems.

Specifying BIT Performance

One of the more complex tasks inherent in the acquisition of modern systems is the development of realistic and meaningful operational requirements and their subsequent conversion into understandable and achievable contractual specifications. This situation is equally applicable to BIT. Before discussing this topic in more detail, we will present typical performance measures or figures of merit which are used to specify BIT performance.

Percent Detection. The percent of all faults or failures that the BIT system detects.

Percent Isolation. The percent of detected faults or failures that the system will isolate to a specified level of assembly. For example, the BIT might isolate to one box or to three or less printed circuit boards in a box.

Automatic Fault Isolation Capability (AFIC). The AFIC is the product of percent isolation times percent detection.

AFIC = % detection x % isolation

Percent of False Alarms. The percent of the BIT indicated faults where, in fact, no failure is found to exist.

Percent of False Removals. The percentage of units removed because of BIT indications which are subsequently found to test "good" at the next higher maintenance station.

For each of the above parameters, there is a considerable span of interpretation. For example, does the percent detection refer to failure modes or the percentage of all failures that could potentially occur? Does the detection capability apply across the failure spectrum, i.e., mechanical systems, instrumentation, connections and software, or is its diagnostic capability applicable only to electronic hardware systems?

A major contractual issue relates to the definition of failure. should BIT performance be viewed in terms of "BIT addressable" failures, which normally exclude cable/connector, etc., problems as not contractually chargeable, or in terms of all operationally relevant maintenance actions?

An important consideration relates to exactly what failures BIT can detect. Our BIT system will operate ineffectively if the 80% of detectable failures occur infrequently while the remaining 20% occur with predictable regularity. It, therefore, becomes important to specify BIT performance measures in relation to overall mission availability requirements.

Relative to isolation characteristics, will the BIT isolate failures while the basic system is in an operational mode, or must the basic system be "shut down" to permit the isolation software package to be run? How does this characteristic impact mission requirements? Also, to what "level" will the BIT system isolate failures? Early BIT systems were frequently designed to fault isolate to the module level. this resulted in BIT systems as complex, and frequently less reliable than, the basic system. The current trend is as, and frequently less reliable than, the basic system. The current trend is to isolate to the subsystem or box level based on BIT's ability to detect abnormal output signal patterns. Intermediate and depot level maintenance facilities will frequently use BIT or external diagnostic equipment to isolate to the board or piece-part level.

The percent of false alarms is a difficult parameter to measure accurately because an initial fault detection followed by an analysis indicating that no fault exists can signify several different occurrences, such as:

- The BIT system erroneously detected a fault.
- An intermittent out-of-tolerance condition exists--somewhere.
- A failure exists but cannot be readily reproduced in a maintenance environment.

The percent of false removals can be a more difficult problem to address. False removals may be caused by:

- Incorrect BIT logic.
- Wiring or connection problems which manifest themselves as faulty equipment.
- Improper match of tolerances between the BIT and test equipment at the next higher maintenance level.

The resolution of each type of false alarm and false removal requires a substantially different response. From a logistic viewpoint, false alarms often lead to false removals creating unnecessary demands on supply and maintenance systems. Of Potentially more concern is the fact that false alarms and removals create a lack of confidence in the BIT system to the point where maintenance or operations personnel may ignore certain fault detection indications. Under these conditions, the BIT system in particular and the maintenance concept in general cannot mature nor provide the support required to meet mission requirements.

The specification of BIT performance must be tailored to the specific system under consideration as well as the available funds and, most importantly, the overall mission requirements. This tailoring activity must include a comprehensive definition of BIT capability based upon the figures of merit presented above.

APPENDIX C
AVAILABILITY

INTRODUCTION

Availability is the parameter that translates system reliability and maintainability characteristics into an index of effectiveness. It is based on the question, "Is the equipment available in a working condition when it is needed?" The ability to answer this question for a specific system represents a powerful concept in itself, and there are additional side benefits that result. An important benefit is the ability to use the availability analysis as a platform to support both the establishment of reliability and maintainability parameters and trade-offs between these parameters. As part of our review of availability, we will separate maintainability into its components (preventive and corrective maintenance and administrative and logistics delay times) to determine the impact of these individual elements on overall system availability.

DEFINITION OF AVAILABILITY

Availability is defined as a measure of the degree to which an item is in an operable and committable state at the start of a mission when the mission is called for at a random point in time.

ELEMENTS OF AVAILABILITY

As is evident by its very nature, approaches to availability are time-related. Figure C-1 illustrates the breakdown of total equipment time into those time-based elements on which the analysis of availability is based. Note that the time designated as "off time" does not apply to availability analyses because during this time system operation is not required. Storage and transportation. periods are examples of "off time."

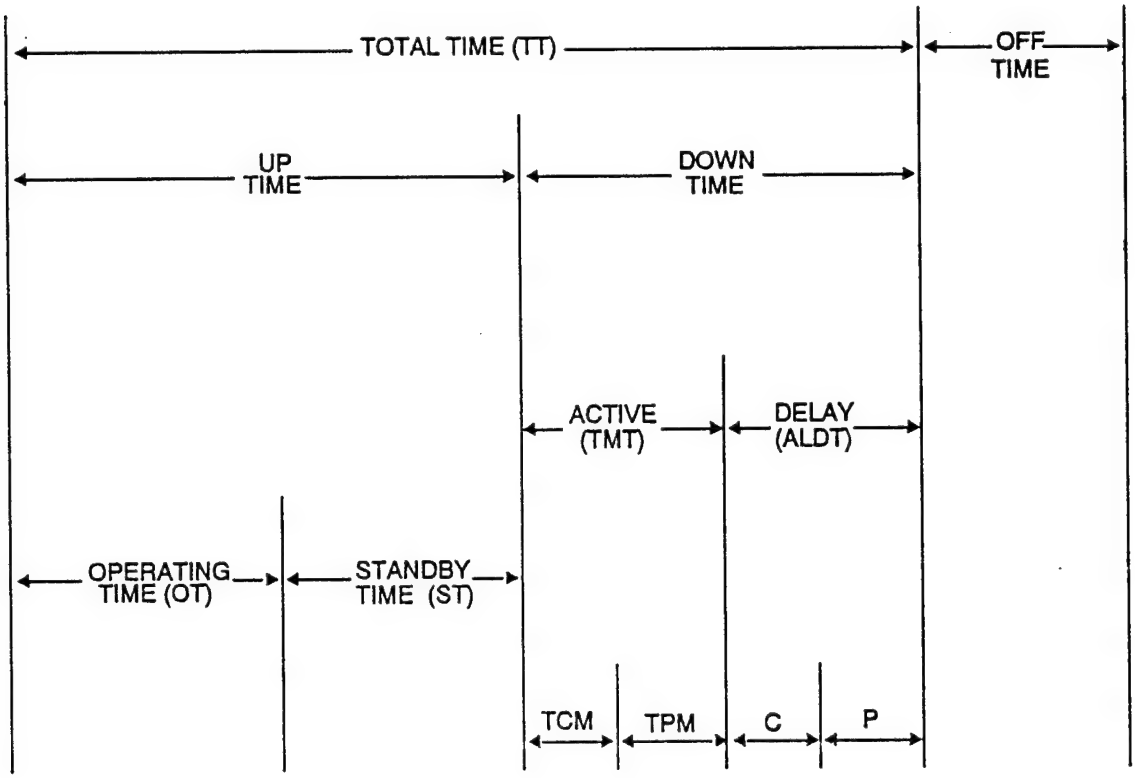


FIGURE C-1. BREAKDOWN OF TOTAL EQUIPMENT TIME

The letters "C" and "P" represent those periods of time attributed to corrective or preventive maintenance, respectively, which are expended in active repair of hardware or in waiting (delay) for resources to effect needed repairs. This waiting or delay period can further be subdivided into administrative and logistics delay periods.

DEFINITION OF TERMS

Definitions of commonly used availability elements are given below. Several are displayed pictorially in Figure C-1.

TT = Total intended utilization period, total time.

TCM = Total corrective (unscheduled) maintenance time per specified period.

TPM = Total preventive (scheduled maintenance time per specified period.

ALDT = Administrative and logistics down time spent waiting for parts,

administrative processing, maintenance personnel, or transportation per specified period. See Figure C-1, Delay-Down Time (no maintenance time).

TMT = Total maintenance time = TCM + TPM. See Figure C-1, Active-Down Time.

TDT = Total down time = TMT + ALDT.

OT = Operating time (equipment in use). See Figure C-1.

ST = Standby time (not operating but assumed operable) in a specified period. See Figure C-1.

MTBF = Mean time between failures.

MTBM = Mean time between maintenance actions.

MTBUMA = Mean time between unscheduled maintenance actions (unscheduled means corrective).

MDT = Mean down time.

MTTR = Mean time to repair.

MATHEMATICAL EXPRESSIONS OF AVAILABILITY

The basic mathematical definition of availability is

$$\text{Availability} = A = \frac{\text{Up Time}}{\text{Total Time}} = \frac{\text{Up Time}}{\text{Up Time + Down Time}}$$

Actual assessment of availability is accomplished by substituting the time-based elements defined above into various forms of this basic equation. Different combinations of elements combine to formulate different definitions of availability.

Operational availability is the most desirable form of availability to be used in assessing a system's combat potential. Achieved, and to a lesser degree inherent availability are primarily the concern of the developing agency in its interface with the contractor and other co-developing agencies.

Ao is an important measure of system effectiveness because it relates system hardware, support and environment characteristics into one meaningful parameter--a figure of merit depicting the

equipment state at the start of a mission. Because it is an effectiveness-related index, availability is used as a starting point for nearly all effectiveness and force sizing analyses.

Inherent Availability (Ai)

Under certain conditions, it is necessary to define system availability with respect only to operating time and corrective maintenance. Availability defined in this manner is called inherent availability (Ai).

$$Ai = \frac{MTBF}{MTBF + MTTR}$$

Under these idealized conditions, we choose to ignore standby and delay times associated with scheduled or preventive maintenance, as well as administrative and logistics down time. Because only corrective maintenance is considered in the calculation, the MTBF becomes MTBUMA, and, likewise, MTTR is calculated using only times associated with corrective maintenance.

Inherent availability is useful in determining basic system operational characteristics under conditions which might include testing in a contractor's facility or other controlled test environment. Likewise, inherent availability becomes a useful term to describe combined reliability and maintainability characteristics or to define one in terms of the other during early conceptual phases of a program when, generally, these terms cannot be defined individually. Since this definition of availability is easily measured, it is frequently used as a contract-specified requirement.

As is obvious from this definition, inherent availability provides a very poor estimate of true combat potential for most systems, because it provides no indication of the time required to obtain required field support. This term should normally not be used to support an operational assessment.

Operational Availability

Operational availability, unlike inherent availability, covers all segments of time that the equipment is intended to be operational (total time in Figure C-1). The same up-down time relationship exists but has been expanded. Up time now includes operating time plus nonoperating (stand-by) time (when the equipment is assumed to be operable). Down time has been expanded to include preventive and corrective maintenance and associated administrative and logistics lead time. All are measure in clock time.

$$\text{Operational Availability} = Ao = \frac{OT + ST}{OT + ST + TPM + TCM + ALDT}$$

This relationship is intended to provide a realistic measure of equipment availability when the equipment is deployed and functioning in a combat environment. Operational availability is used to support operational testing assessment, life cycle costing, and force development exercises.

One significant problem associated with determining Ao is that it becomes costly and time-consuming to define the various parameters. Defining ALDT and TPM under combat conditions is not feasible in most instances. Nevertheless, the operational availability expression does provide an accepted technique of relating standard reliability and maintainability elements into an effectiveness-oriented parameter. As such, it is a useful assessment tool.

One important aspect to take note of when assessing Ao is that it is affected by utilization rate. The less a system is operated in a given period, the higher Ao will be. It is important therefore when defining the "total time: period to exclude lengthy periods during which little or no system usage is anticipated.

One other frequently encountered expression for operational availability is

$$Ao = \frac{OT}{MTBM + MDT}$$

where

MTBM = mean time between maintenance actions and MDT = mean down time.

While maintenance-oriented, this form of Ao retains consideration of the same basic elements. The MDT interval includes corrective and preventive maintenance and administrative and logistics down time. This form of the AO relationship would generally prove more useful in support of early maintenance parameter sensitivity and definition analysis. Note that the above definition assumes that standby time is zero.

Achieved Availability (Aa)

This definition of availability is mathematically expressed as

$$Aa = \frac{OT}{OT + TCM + TPM}$$

Aa is frequently used during development testing and initial production testing when the system is not operating in its intended support environment. Excluded are operator before-and-after maintenance checks and standby, supply, and administrative waiting periods. Aa is much more a system hardware-oriented measure than is operational availability, which considers operating environment factors. It is, however, dependent on the preventive maintenance policy, which is greatly influenced by non-hardware considerations.

A GENERAL APPROACH FOR EVALUATING AVAILABILITY

The following paragraphs present a generalized approach for evaluating system availability. It is important to note that for such an analysis to be meaningful to an equipment user or developer it must reflect the peculiarities of the system being considered.

General Procedure

1. The operational and maintenance concepts associated with system utilization must be defined in detail using terminology compatible with the user, developer an d contractor.

2. Using the above definitions, construct a time line availability model (see Figure C-2) which reflects the mission availability parameters.

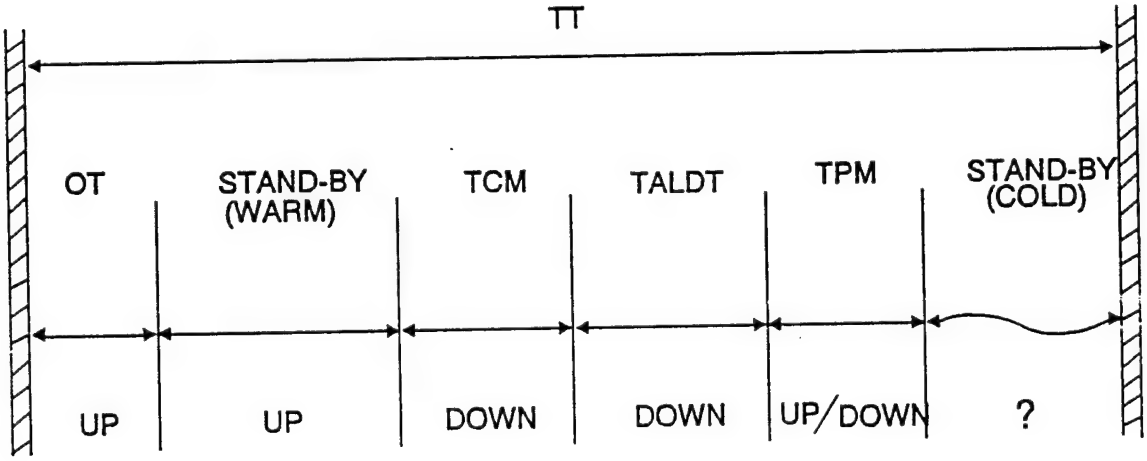


FIGURE C-2. MISSION AVAILABILITY TIME LINE MODEL GENERALIZED FORMAT

NOTE: Figure C-2 displays elements of availability frequently included in a quantitative assessment of availability. The up or down status of a specific system during preventive maintenance must be closely examined. Generally, a portion of the preventive maintenance period may be considered as uptime. Cold standby time must also be examined closely before determining system up or down status during this period.

3. With the aid of the time line model, determine which time elements represent "uptime" and "downtime." Don't be mislead by the apparent simplicity of this task. For example, consider

that the maintenance concept may be defined so that the equipment must be maintained in a committable state during the performance of preventive maintenance.

Additionally, for multi-mission and/or multi-mode systems, it will be necessary to determine up and down times as a function of each mission/mode. This generally will require the use of a separate time line model for each identifiable mission/mode.

Likewise, separate time line models are generally required to support the availability analyses of systems which experience significantly different peacetime, sustained combat and surge utilization rates.

4. Determine quantitative values for the individual time elements of the time line models. Coordinate these values with the user, developer and contractor.

5. Compute and track availability using the definitions of availability appropriate for the current stage of system development.

6. Continue to check availability model status and update the model as required. Special attention should be given to updating the model as the operational, maintenance, and logistics support concepts mature.

System Availability Assessment Considerations

As indicated in the above paragraphs, the quantitative evaluation of availability must be carefully and accurately tailored to each system. For this reason, no detailed examples are presented in this text. However, the following paragraphs do present concepts which will apply to various classes of systems.

Recovery Time

Normally, availability measures imply that every hour has equal value from the standpoint of operations and the performance of maintenance and logistics activities. Normally, the operational concept requires the system to function only for selected periods. The remaining time is traditionally referred to as "off-time," during which no activity is conducted.

An alternative to the "off-time" or "cold standby" concepts is the use of the term "recovery time' (RT).

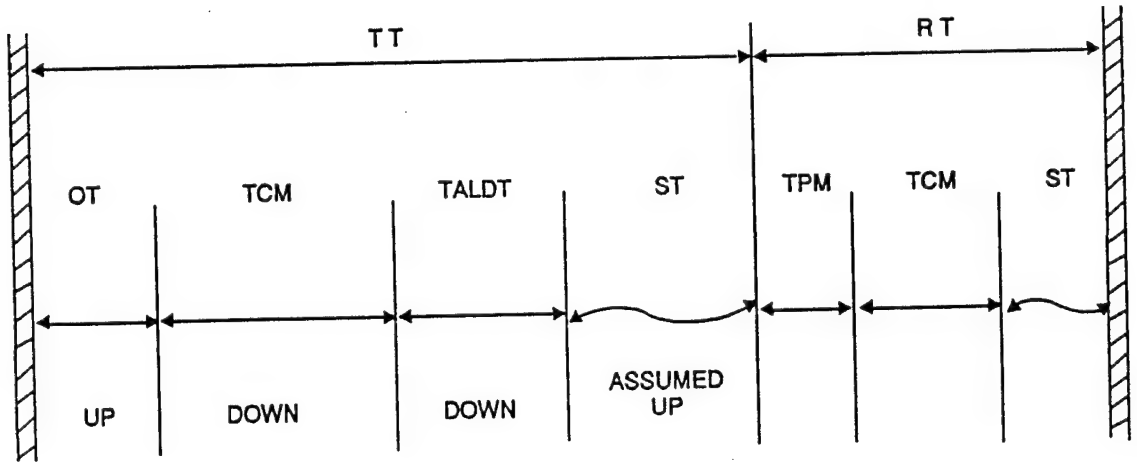


FIGURE C-3. MISSION AVAILABILITY TIME LINE MODEL RECOVERY TIME FORMAT

Recovery time represents an interval of time during which the system may be up or down. Recovery time does not appear in the availability calculation which based only on the TT time period. Take special note of the fact that corrective maintenance time (TCM) is found in both TT and RT time intervals. Corrective maintenance performed during the TT period is maintenance required to keep the system in a mission ready or available status. Corrective maintenance performed during the RT period generally addresses hardware malfunctions which do not result in a non mission-ready status.

The principal advantage of using the "recovery time" analysis is that it can provide a more meaningful availability assessment for systems whose periods of required availability are predictable and whose preventive maintenance constitutes a significant but delayable portion of the maintenance burden.

The recovery time calculation technique concentrates the availability calculation during the operational time period, thereby focusing attention on critical up and down time elements.

The above discussion presents an alternate technique of computing system availability, i.e., the use of the recovery time concept. Whatever technique is selected for computing availability, it must be carefully tailored to the system undergoing assessment.

Definition of the terms used in availability analysis must be stressed. For example, what total time (TT) period has been chosen for an analysis base? Assume for a moment that we are

assessing the Ao of an operational squadron and that we have chosen a 7-day TT period. If the aircraft normally are not flown on weekends or are left in an up condition on Friday night it is obvious that Ao will be higher than if a 5-day total time were selected. Reference the discussion of recovery and standby time.

Other definitions associated with Ao are not quite so obvious and must be included in pretest definition. For example, are "before and after" operational checks conducted in conjunction with preventive maintenance excluded from down time because the equipment is assumed operable? Similarly, are corrective maintenance diagnostic procedures logged against down time? What if the a hardware is not found defective? How is ALDT arrived at? Is it assumed, calculated or observed? What is the operational status of a system during the warm standby period?

HARDWARE REQUIREMENT ESTABLISHMENT AND TRADE-OFFS

The expression for availability frequently provides the vehicle needed to analyze other system requirements both directly and by way of trade-offs.

AVAILABILITY FOR MULTI-MISSION SYSTEMS

For many modern weapon systems, availability is not simply an "up" or "down" condition. Systems such as AEGIS and HAWK have multi-mission/mode capabilities and thus require detailed techniques to characterize the associated availability states. While these multi-mission/mode characterizations may appear different, they are indeed based on the expressions presented previously. The definition of terms, modes and states is equally important in the analysis of these complex systems.

SIMULATION MODELS

There are a number of computer simulation models available which are well suited for evaluating interactions between system design characteristics, logistic support, and relevant operational output measures such as operational availability or sortie rate. Examples of such models include LCOM (aircraft), CASEE and PRISM (carrier-based aircraft), ARMS (Army aircraft), TIGER (Ship systems), RETCOM (combat vehicles), etc. These models provide visibility of manpower and test equipment, queuing effects, and the impact of spares stockage levels on operational availability, which generally cannot be evaluated with simple analytical formulas. Simulation models are particularly useful for suing test results to project operational availability under conditions different from the test environment (e.g., to project availability under different from the test environment (e.g., to project availability under wartime surge conditions). One drawback to simulation models is that they are usually more cumbersome to use than straightforward analytical techniques.

APPENDIX D
STATISTICAL CONCEPTS

INTRODUCTION

Our assumptions about a given testing situation lead us to the choice of a mathematical model to characterize the reliability of a system. However, we cannot determine the actual reliability of the system using the model until the parameters of the model, p for the binomial and λ (or θ) for the Poisson or exponential model, have been specified. The values of the parameters are never known with absolute certainty. As a consequence, some form of sampling or testing is required to obtain estimates for these parameters. The quality of the estimates is, of course, directly related to the quality and size of the sample.

POINT ESTIMATES

Point estimates represent a single "best guess" about model parameters, based on the sample data. A distinguishing symbol commonly is used to designate the estimate of a parameter. Most commonly is used to designate the estimate of a parameter. Most commonly, a caret or "hat" is used to designate point estimates (e.g., $\hat{\theta}$, $\hat{R}(x)$, $\hat{\lambda}$). Quite often, and for our purposes, the caret further indicates that the estimator is a maximum likelihood estimator; that is, it is the most likely value of the parameter of the model which is presumed to have generated the actual data.

There are criteria other than maximum likelihood used for a single "best guess." One other is unbiasedness. For an estimator to be unbiased, we mean that, in the long run, it will have no tendency toward estimating either too high or low. The point estimates which we propose for p in the binomial model and for λ in the Poisson and exponential models are both maximum likelihood and unbiased.

CONFIDENCE STATEMENTS

Point estimates represent a single "best guess" about parameters, biased on a single sample. The actual computed values could greatly overestimate or underestimate the true reliability parameters, particularly if they are based on a small amount of data. As an example, suppose that 20 rounds of ammunition were tested and 18 fired successfully.

The maximum likelihood and unbiased estimate of reliability is $\hat{R} = 18/20 = 0.9$. In other words, the system most likely to have generated 18 successes is one whose reliability is 0.9. Note that 0.9 is the percentage of successes actually observed in the sample. However, a system whose true reliability is somewhat less than or somewhat more than 0.9 could reasonably have generated this particular data set.

We use confidence limits to address how high or low the value of a parameter could reasonably be. a 90% confidence interval for reliability is: $0.717 < R < 0.982$. In other words, if being reasonable signifies being 90% confident of being right, then it is unreasonable to consider that a system whose reliability is actually less than 0.717 or one whose reliability is actually more than 0.982 generated the 18 successful rounds. When we desire to be more confident, say 95% confident, that our interval contains the true system reliability, we widen our interval, i.e., we expand the group of systems considered to have reasonably generated the data. A 95% confidence interval for the reliability of our example system is: $0.683 < R < 0.988$. Since we are now allowing for the possibility that the system reliability could be a little lower than 0.717 -- namely, as low as 0.683 -- or a little higher than 0.988 - we can now afford to be more confident that our interval indeed contains the true value. For a fixed amount of testing, we can only increase our confidence by widening the interval of reasonable values.

Suppose that we desire to reduce the size of the interval while maintaining the same level of confidence or to increase the level of confidence while maintaining approximately the same size interval. Either of these objectives is accomplished through increased testing, i.e., taking a larger sample. If the system test had resulted in 27 successful firings out of 30 attempts (vice 18 out of 20), the point estimate is still 0.9. However, the 90% confidence interval for system reliability is: $0.761 < R < 0.972$. The length of this interval represents a 20% reduction in the length of the 90% confidence interval resulting from our test of 20 units. The 95% confidence interval for system reliability is: $0.734 < R < 0.979$. This interval represents an 8% reduction in size, but our confidence has increased to 95%. Figure D-1 graphically portrays the effect on interval length induced by changing confidence levels or increasing sample size.

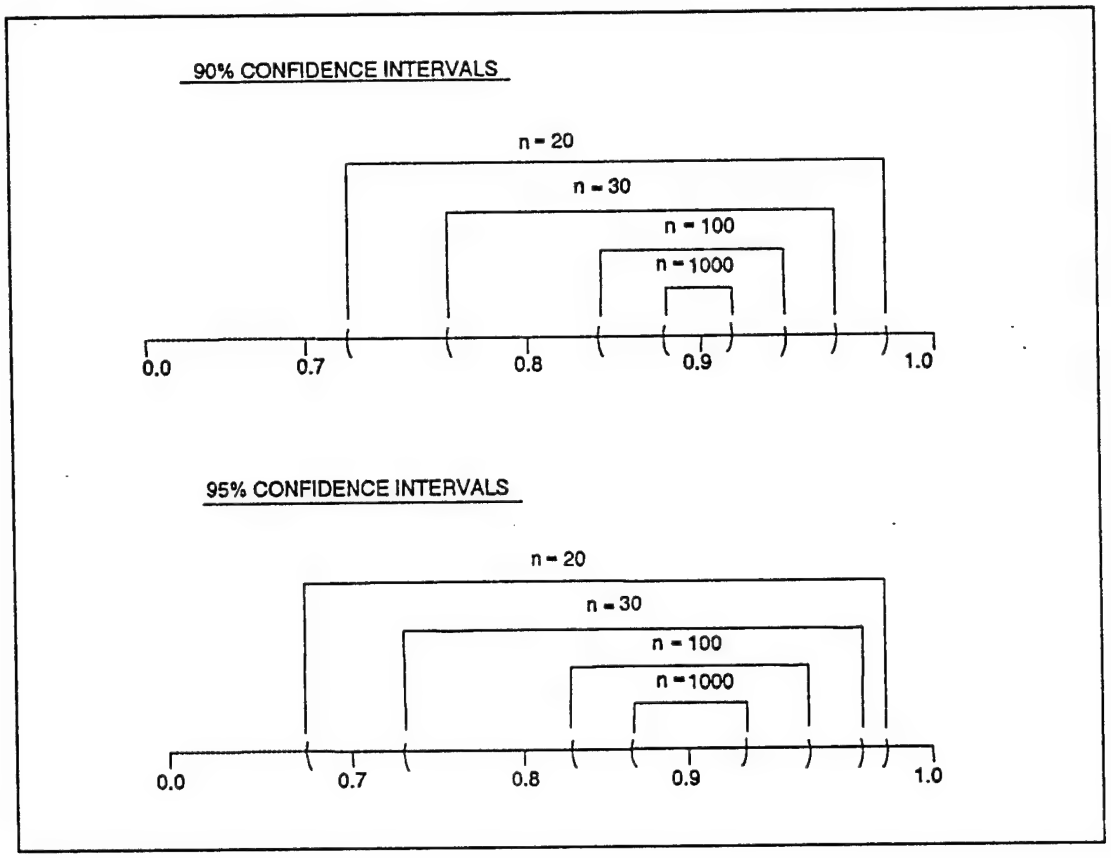


FIGURE D-1 CONFIDENCE INTERVALS

A cautious, conservative person who buys safe investments, wears a belt and suspenders, and qualifies his statements carefully is operating on a high-confidence level. He is certain he won't be wrong very often. If he is wrong once in 100 times, he is operating on a 99% confidence level. A less conservative person who takes more chances will be wrong more often, and hence he operates on a lower confidence level. If he is wrong once in 20 times, he is operating on a 95% confidence level. The confidence level, therefore, merely specifies the percentage of the statements that a person expects to be correct. If the percentage of the statements that a person expects to be correct. If the experimenter selects a confidence level that is too high, the test program will be prohibitively expensive before any very precise conclusions are reached. If the confidence level is too low, precise conclusions will be reached easily, but these conclusions will be wrong too frequently, and, in turn, too expensive if a large quantity of the item is made on the basis of erroneous conclusions. There is no ready answer to this dilemma.

We can interpret confidence statements using the concept of risk. With a 90% confidence statement, there is a 10% risk; with a 99% confidence statement, there is a 1% risk. Confidence

intervals generally are constructed so that half of the total risk is associated with each limit or extreme of the interval. Using this approach with a 90% interval for reliability, there is a 5% risk that the true reliability is above the upper limit. We can therefore state for the example system with 18 of 20 successes that we are 95% confident that: $R > 0.717$. This is a lower confidence limit statement. We are also 95% confident that: $R < 0.092$. This is a upper confidence limit statement. See figure D-2.

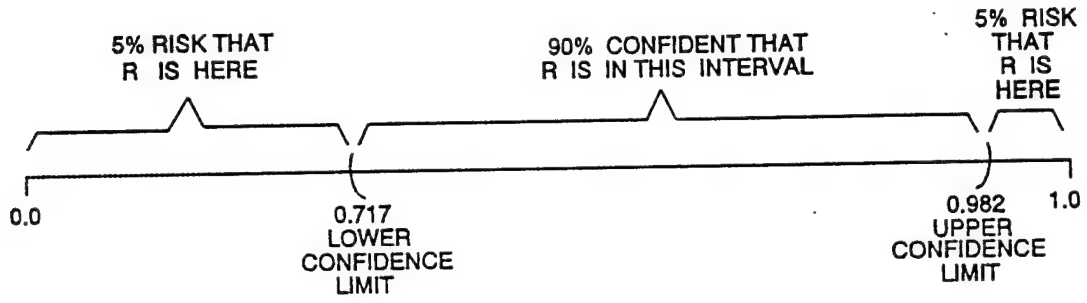


FIGURE D-2. UPPER CONFIDENCE LIMITS

The classical textbook approach to confidence intervals has been to specify the desired confidence level and determine the limit associated with this confidence level. This approach creates a twofold problem. First, the desired confidence level has to be determined. Second, the limits that are generated are generally not, in themselves, values of direct interest. A very practical modification is to determine the level of confidence associated with a predetermined limit value. For example, the minimum value of a reliability measure that is acceptable to the user is a logical lower limit. the confidence in this value can then be interpreted as the assurance that the user's needs are met. See Figure D-3.

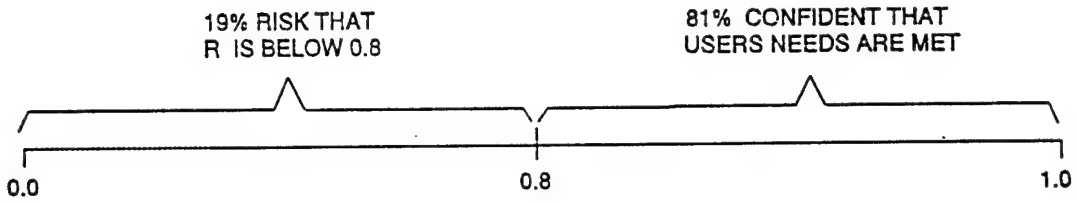


FIGURE D-3. CONFIDENCE INTERVALS - ACCEPTABLE LOWER LIMITS

The confidence level for a lower limit of 0.8 is 81%. A system reliability of 0.8 is the user's minimum acceptable value (MAV).

HYPOTHESIS TESTING

While confidence limits are generally used to define the uncertainty of a parameter value, an alternative approach is hypothesis testing. Both approaches essentially give the same information. Hypothesis testing can be used to distinguish between two values or two sets of values for the proportion of failures in a binomial experiment, or for the failure rate in a Poisson/exponential experiment. Let us examine hypothesis testing using a binomial example. Typically, for a binomial experiment, it is hypothesized that the probability of failure, p, is a specified value. While there is seldom any belief that p is actually equal to that value, there are values of p which would be considered unacceptable in a development program. These unacceptable values are specified in an alternative hypothesis. Consider the following examples.

(1) One-Sided Tests

H_o: $p = 0.3$ (Null Hypothesis)

H_1: $p > 0.3$ (Alternative Hypothesis)

In Case (1), the evaluator hopes that p is no more than 0.3. He considers a p of more than 0.3 to be unacceptable. This is a classical one-sided test. Another type of one-sided test has the alternative hypothesis $p < 0.3$.

(2) Two-Sided Tests

H_o: $p = 0.3$

H_1: $p \neq 0.3$

In Case (2), the evaluator hopes that p is approximately 0.3. Values of p much larger than or much smaller than 0.3 are unacceptable. This is a classical two-sided test.

(2) Simple vs. Simple Tests

H_o: $p = 0.3$

H_1: $p = 0.5$

In case 3, the evaluator hopes that p is no more than 0.3. He considers a p of more than 0.5 to be unacceptable. The region between 0.3 and 0.5 is an indifference region in that it represents acceptable but not hoped for values. This is actually a classical simple versus simple test.

In order to conduct a statistical test of hypothesis, the following steps are employed:

1. The hypothesis, null and alternative, are specified. For our purposes, the null hypothesis is the contractually specified value (SV) and the alternative hypothesis is the minimum acceptable value (MAV).

2. A sample size, n, is determined. This value must be large enough to allow us to distinguish between the SV and MAV.

3. An accept/reject criterion is established. For our purposes, this criterion is established by specifying a value c, which is the maximum number of failures permitted before a system will be rejected.

4. The sample is taken and the hypothesis is chosen based upon the accept/reject criterion. If c or fewer failures occur, we accept the system. If more than c failures occur, we reject the system.

PRODUCER'S AND CONSUMER'S RISKS

There are two possible errors in making a hypothesis-testing decision. We can choose the alternative hypothesis, thereby rejecting the null hypothesis, when, in fact, the null hypothesis is true. The chance or probability of this occurring is called the producer's risk, α. On the other hand, we can choose the null hypothesis, i.e., accept it as reasonable, when in fact the alternative hypothesis is true. the chance or probability of this occurring is termed the consumer's risk, β.

Consider the following: A system is under development. It is desired that i have a 300-hour MTBF. However, an MTBF of less than 150 hours is unacceptable, i.e., the MAV is 150 hours. How would we set up a hypothesis test to determine the acceptability of this new system? Our null hypothesis (desired value) is that the MTBF is 300 hours. Our alternative hypothesis (values of interest) is that the MTBF has a value which is less than 150 hours. To decide which hypothesis we will choose, we determine a test exposure and a decision criterion. The α risk (producer's risk) is the probability that the decision criterion will lead to a rejection decision when in fact the system meets the specification of 300 hours MTBF. The β risk (consumer's risk) is the probability that the decision criterion will lead to an acceptance decision when in fact the system falls short of the 150 hours MTBF.

For a given test, the decision criteria can be altered to change the α and β risks. Unfortunately, a decision criterion which decreases one automatically increases the other. The only way to decrease both risks is to increase the test exposure, that is, the number of test hours.

In both test planning and data analysis situations, either hypothesis testing or confidence statements provide an avenue of approach. The interface between the two approaches can be best understood through the following example.

Suppose α is the desired producer's risk (A1 = 0.05) for the specified MTBF of 300 hours. Suppose further that α is the desired consumer's risk (α = 0.1) for the minimum acceptable MTBF of 150 hours. The hypothesis testing approach determines a required sample size and a specified accept/reject criterion. We show how the same information can be obtained through confidence statements in the following two cases. The abbreviations LCL and UCL represent Lower Confidence Limit and Upper Confidence Limit, respectively.

Note that the distance between the upper and lower limits is the same as the distance between the SV and the MAV. When this is the case we shall always be able to make a clear-cut decision and the risks associated with the decision will be as specified at the outset of testing.

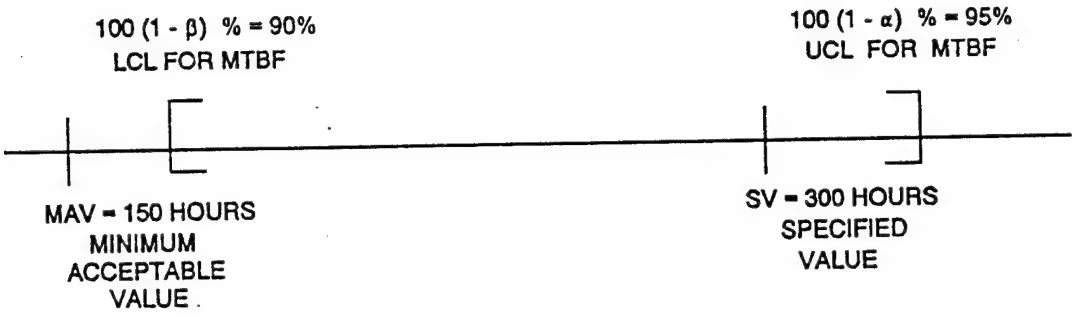


FIGURE D-4. ACCEPTANCE DECISION

Note that in Figure D-4 the 100(1-β)% = 90% lower limit exceeds the MAV of 150 hours. In addition, the 100(1-α)% = 95% upper limit exceeds the specified value of 300 hours. The consumer is 90% confident that the 150-hour MAV has been met or exceeded and the producer has demonstrated that the system could reasonably have a 300-hour MTBF. consequently, we would make the decision to accept the system.

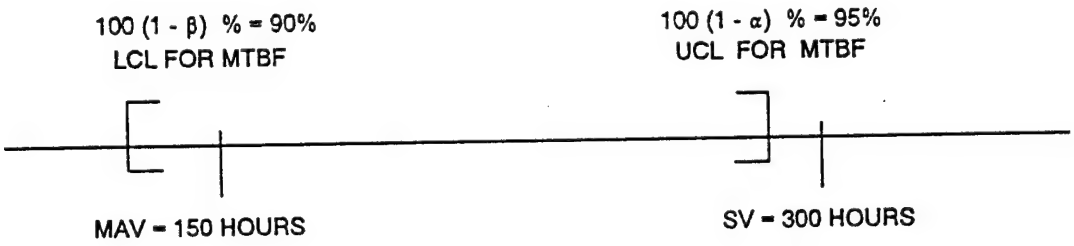


FIGURE D-5. REJECTION DECISION

Note that in Figure D-5 the 100(1-β)% = 90% lower limit falls below the MAV of 150 hours. In addition, the 100(1-α)% = 95% upper limit falls below the SV of 300 hours. Therefore, the true MTBF could reasonably be below 150 hours and the producer has not demonstrated that an MTBF of 300 hours is reasonable. Consequently, we make the decision to reject the system.

TEST EXPOSURE

Perhaps one of the most important subjects to be considered in the evaluation of RAM characteristics is the subject of test exposure. The term "test exposure" refers to the amount (quantity and quality) of testing performed on a system or systems in an effort to evaluate performance factors.

Recall the comment we made in the previous section to the effect that the difference in the distance between the upper and lower confidence limits was equal to the difference in the distance between the SV and the MAV. When this condition is achieved, we have obtained the most efficient test exposure for the stated requirements and risks. Examples of situations where test exposure is inadequate or excessive are given below.

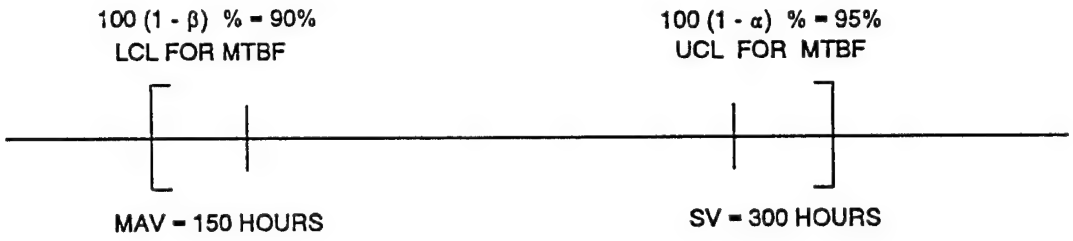


FIGURE D-6. INADEQUATE TEST DURATION

Note that in Figure D-6 the $100(1-\beta)\% = 90\%$ lower limit falls below the MAV of 150 hours. The $100(1-\alpha)\% = 95\%$ upper limit exceeds the SV of 300 hours. the true MTBF could reasonably be below 150 hours or above 300 hours. Test exposure is insufficient to discriminate between the MAV of 150 hours and the SV of 300 hours with the required risk levels of 10% and 5%. If we reject the system, the producer can legitimately claim that an MTBF of 300 hours is reasonable for his system. On the other hand, if we accept the system, we may be fielding an inadequate system.

Note that in Figure D-7 the $100(1-\beta)\% = 90\%$ lower limit exceeds the MAV of 150 hours. The $100(1-\alpha)\% = 95\%$ upper limit falls below the SV of 300 hours. The consumer has 90% confidence that the 150-hour MAV has been met or exceeded. However, the producer has not demonstrated the specified 300-hour MTBF. The test exposure is more than required to obtain the risks of 10% and 5% for the stated values of MAV and SV. Since the MAV has been met or exceeded, we will probably accept the system. We may have paid a premium to obtain information that allowed us to construct a confidence interval more narrow than required.

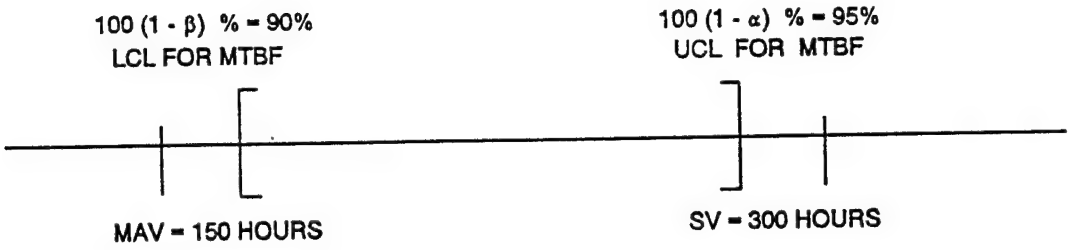


FIGURE D-7. EXCESSIVE TEST DURATION

APPENDIX E
RELIABILITY TEST PLANNING

INTRODUCTION

This appendix presents the techniques for determining the amount of test exposure required to satisfy previously established program reliability requirements. The reader will note that Appendix D addresses the topic of reliability data analysis. There, we assumed that the test data had already been gathered. We then used the available data to determine point estimates for reliability parameters and to stipulate the uncertainty associated with these estimates.

Appendix E presents techniques for designing test plans which can verify that previously specified reliability requirements have been achieved. We realize, of course, that the required test exposure and/or sample size may exceed the available resources. In such cases, alternative test plans, consistent with program constraints, must be developed. In this appendix, we also present methods which make it possible to clearly identify the inherent risks associated with a limited test program.

PRIMARY TEST DESIGN PARAMETERS

Upper and Lower Test Values

Two values of system reliability are of particular importance in the design of a reliability test plan. These are referred to as the upper test and lower test values. In some cases, only a single value is initially apparent, the second value being only implied. These two values and the risks associated with them determine the type and magnitude of testing required.

The upper test value is the hoped for value of the reliability measure, An upper test MTBF is symbolized as θ_0, and an upper test reliability is symbolized as R_0. A test plan is designed so that test systems whose true reliability parameters exceed θ_0 and R_0 will, with high probability, perform during the test in such a way as to be "accepted."

The lower test value is commonly interpreted in two different ways that may initially appear contradictory. One interpretation is that this lower value of the reliability measure represents a rejection limit. The other interpretation is that this value is minimally acceptable. The apparent conflict is resolved by viewing the lower test value as the fine line between the best rejectable value and the worst acceptable value. A lower test MTBF is symbolized as θ_1, and a lower test reliability is symbolized as R_1. Systems whose true reliability parameters having values less than θ_1 and R_1 will, with high probability, perform in such a way as to be "rejected."

The upper and lower test values serve to divide the reliability, or MTBF, scale into three distinct regions as shown in Figure E-1. Note that the region between R_1 and R_0 is neither bad enough to demand rejection nor is it good enough to demand acceptance. This region is necessary since we will never precisely know the true reliability of the system.

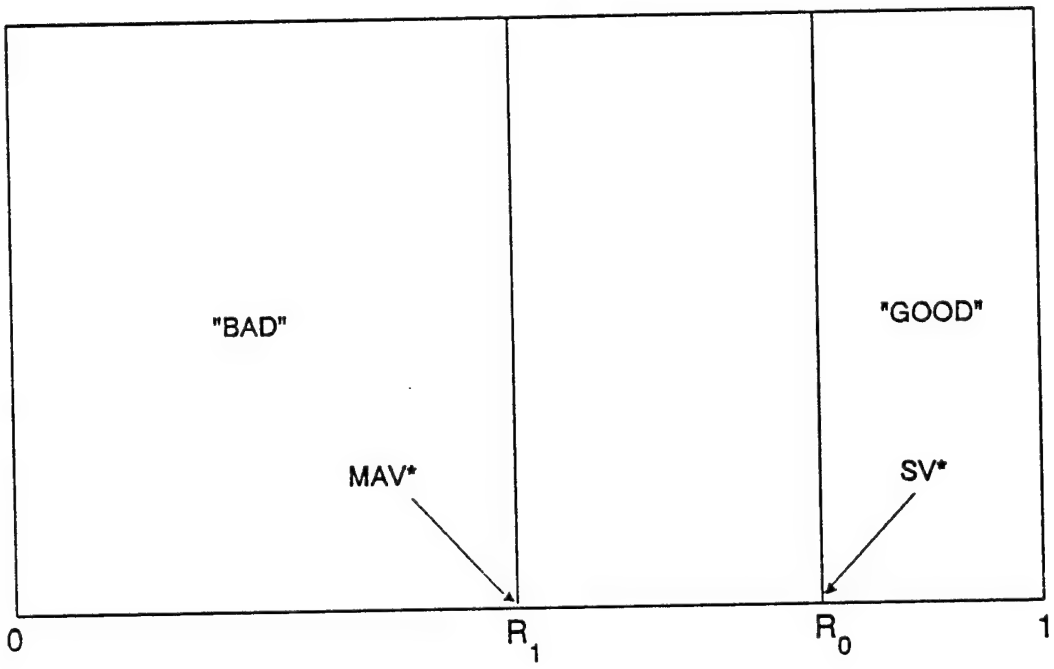


FIGURE E-1. REGIONS DEFINED BY R_0 AND R_1

The user's reliability requirement should be stated as a minimum acceptable value (MAV); that is, the worst level of reliability that the user can tolerate and accept. The contractually specified value (SV) is a value somewhat higher than the MAV. For reliability qualification tests prior to production, the lower test value is the MAV, and the upper test value is the SV. Earlier in the development process, fixed configuration test may be conducted to demonstrate the attainment of lower levels of reliability at specified milestones. In such cases, upper test and lower test values should be consistent with the stage of the development process.

In the above paragraphs, we have been discussing population parameter values only. These values are never known with absolute certainty, so we are forced to base an evaluation of system performance characteristics on sample data. Let us conclude this section with a discussion of

sample reliability values and how we can interpret them to aid us in making our system reliability assessment.

One objective of this appendix is the determination of an accept/reject criterion for a test to be conducted. As an example, consider the value R_T in Figure E-2 below. The term R_T is that value of the sample reliability which corresponds to the maximum allowable number of failures that can occur during testing and still result in acceptance of the system.

If we test our determined number of articles and find that R_{sample} is larger than R_T, then we accept the system because there is high probability that the

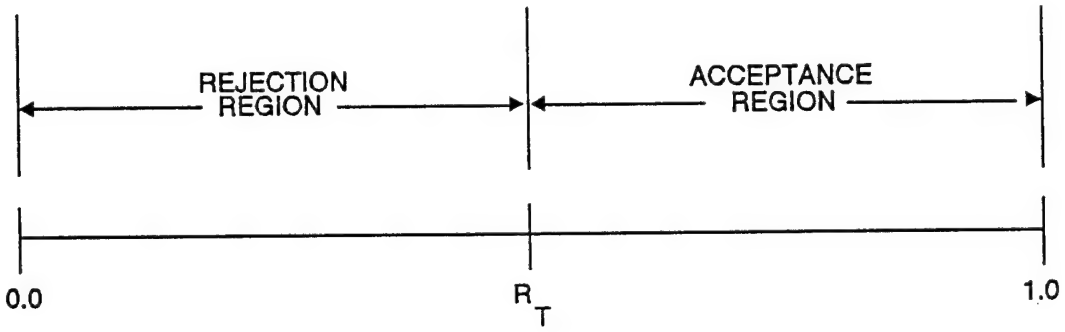


FIGURE E-2. SAMPLE RELIABILITY RANGE

sample system (s) come from a population of systems whose true reliability R exceeds R_1, the minimum acceptable value (MAV) (see Figure E-1) for this test. Note that when R_{sample} is larger than R_T, we have confidence that the true reliability exceeds the MAV. We should not interpret this result as an indication that the contractor has met the SV. Further, if R_{sample} is smaller than RT, we will reject the system because there is high probability that the sample system(s) come from a population whose true reliability R is lower than R_O, the SV for this test. Note that when R_{sample} is smaller than R_T, we have confidence that the true reliability falls short of the SV. We should not interpret this result as an indication that the MAV has not been met, but rather that the MAV has not been demonstrated at a sufficiently high level of confidence.

Consumer's Risk (β) and Producer's Risk (α)

The consumer's risk (β) is the probability of accepting the system if the true value of the system reliability measure is less than the lower test value. It can be interpreted in the following ways:

1. β represents the maximum risk that the true value of the reliability measure is, in fact, less than the lower test value.

2. From an alternative viewpoint, if the acceptance criterion is met, there will be at least $100(1-\beta)\%$ confidence that the true value of the reliability measure equals or exceeds the lower test value.

The producer's risk (α) is the probability of rejection if the true value of the reliability measure is greater than the upper test value. It can be interpreted in the following ways:

1. The probability of acceptance will be at least $(1-\alpha)$ if the upper test value is, in fact, met or exceeded.

2. From an alternative viewpoint, if there is a rejection decision, there will be at least $100(1-\alpha)\%$ confidence that the true value of the reliability measure is less than the upper test value.

Pre- and Post-Test Risk Considerations

Before proceeding on with the application of the consumer's and producer's risk concept, it is important to understand the contrast that exists between pre- and post-test risks.

The α and β risks represent uncertainties that exist in the test planning or pre-data environment discussed in this chapter. Once data has been gathered and we have accepted or rejected the system, we find that the risk environment is altered. For example, we take a sample and decide to accept the system. At this point the producer's risk is eliminated; the consumer's risk remains but is less than the maximum that would exist had the sample reliability, R_{sample}, been exactly equal to R_T.

If, on the other hand, R_{sample} is less than R_T, i.e., we reject the system, we find that the consumer's risk is eliminated since there is no risk of accepting a bad system. Likewise, the producer's risk is less than the maximum that would exist had the sample reliability R_{sample} been exactly equal to R_T.

In this chapter, we are concerned with pre-test risks. We determine the maximum α and β risks and then calculate the required test exposure and acceptable number of failures which will limit our risk to the maximum levels.

TEST DESIGN FOR DISCRETE TIME TESTING: BINOMIAL MODEL

Four values specify the plan for a binomial test. They are:

- the specified or desired proportion of failures (p_o),

- the maximum acceptable proportion of failures (p_1),
- the consumer's risk (β),
- the producer's risk (α).

The test plan itself consists of a sample size (n) and an acceptance criterion (c). The value c represents the maximum number of failures which still results in acceptance of the system. It is usually not possible to construct a plan which attains the exact values of α and β. We shall present methods for determining these types of plans, though in a real world situation, the user and producer may trade off some protection to achieve other goals.

The following paragraphs present exact and approximate procedures to be used in planning a Discrete Time-Binomial Model test program. The "exact procedure" presents the equations used to determine the two values required to specify a binomial test plan. These equations are presented here for the sake of completeness. The "approximate solution" procedure, which makes use of the binomial tables to simplify the procedure, is intended for use by our readers.

Exact Solution Procedure

The exact procedure for determining test plans for the four values listed above is to solve the following two inequalities simultaneously for c and n.

$$\sum_{k=0}^{c} \binom{n}{k} p_1^k (1-p_1)^{n-k} \leq \beta \qquad \text{(E.1)}$$

$$\sum_{k=c+1}^{n} \binom{n}{k} p_o^k (1-p_o)^{n-k} \leq \alpha \qquad \text{(E.2)}$$

There are an infinite number of solutions to this pair of inequalities. The plans of interest are, of course, those which minimize the sample size (n) required. Solving inequalities E.1 and E.2 directly is next to impossible without the aid of a computer. MIL-STD-105D contains numerous binomial test plans which may be used for reliability applications. We should point out that the user unfamiliar with this document will find it difficult to interpret, thus we present the following procedures.

Approximate Solution Procedures

The following so-called approximate procedures utilize the normal and Poisson distributions to obtain approximate solutions to equations E.1 and E.2 and thereby estimate values of the sample size (n) and the acceptance criterion (c). After approximate values for these parameters have been obtained, we may then use the values in conjunction with the binomial tables (Appendix B) and the previously selected and fixed values of α and β to "fine tune" the approximate values of n and c.

Test Planning Using Normal Approximation. The normal distribution provides good approximations for solving inequalities E.1 and E.2, especially for moderate values of p ($0.1 \leq p \leq 0.9$). Using this information, we obtain the approximate solutions for n and c as follows.

$$n = z_\alpha^2 (p_o - p_o^2) + z_\beta^2 (p_1 - p_1^2) + 2 z_\alpha^{z_\beta} \sqrt{p_o p_1 (1-p_o)(1-p_1)} (p_1 - p_o)^2 \qquad \text{(E. 3)}$$

$$c = z_\alpha \sqrt{np_o(1-p_o)} + np_o - 0.5. \qquad \text{(E.4)}$$

Generally, the values computed using equations E.3 and E.4 are good approximations for the test planner. When p_o and p_1 are very small (less than 0.05), the procedure is not recommended. Fine-tuning of the test plan may still require solving the original inequalities or some bargaining with user and/or producer.

As an example, suppose that the minimum acceptable reliability of a system is 0.85 ($p_1 = 0.15$), while the contractually specified reliability is 0.95 ($p_o = 0.05$). Consumer and producer risks of 0.11 are required, i.e., $\alpha = \beta = 0.11$. For $\alpha = 0.11$, $z_\alpha = 1.225$ and for $\beta = 0.11$, $z_\beta = 1.225$. (These values of z_α and z_β are obtained from the normal distribution tables. Using the normal approximation, we have

$$n = \{(1.225)^2 (0.05 - 0.0025) + (1.225)^2 (0.15 - 0.0225)$$

$$+ 2(1.225)^2 \sqrt{(0.05)(0.15)(0.95)(0.85)} / (0.15 - 0.05)^2$$

$$= 49.6 \quad \text{and}$$

$$c = 1.225\sqrt{(49.6)(0.05)(0.95)} + (49.6)(0.05) - 0.5$$

$$= 3.9.$$

The values of n = 49.6 and c = 3.9 are initial approximations. In order to fine tune the test plan, we round these values to n = 50 and c = 4 and use the binomial tables (Appendix B, Table 1). For an n of 50 and a c of 4, the probability of c or fewer failures when $p = p_1 = 0.15$ is 0.1121. In addition, the probability of c or fewer failures when $p = p_o = 0.05$ is 0.8964. Thus, for the test using a sample size of 50 with a maximum acceptable number of failures of 4, the producer's risk $\alpha = 1 - 0.8964 = 0.1036$, and the consumer's risk $\beta = 0.1121$. Note that these values were obtained directly from Binomial tables. It would, however, have been difficult at best to decide where to begin looking in the binomial tables without having first used the normal approximation for guidance.

Test Planning Using Poisson Approximation. The Poisson distribution also provides reasonable approximations to inequalities E.1 and E.2. All this amounts to is substituting np for λt or t/θ in the Poisson distribution equation. Consequently, approximate values for n and c are obtained by solving the following inequalities.

$$\sum_{k=0}^{c} \frac{(np_1)^{k_e} - np_1}{k!} \leq \beta. \qquad (E.5)$$

$$\sum_{k=0}^{c} \frac{(np_0)^{k_e} - np_0}{k!} \geq 1 - \alpha. \qquad (E.6)$$

Standard test plans and procedures for the Poisson (exponential) are readily available and may be used in lieu of solving inequalities E.5 and E.6. This subject is discussed in the "Sources of Exponential Test Plans" section of this chapter. To use these plans in this context, we let $\theta_0 = 1/p_0$, $\theta_1 = 1/p_1$, n = T, and use the acceptable number of failures as given.

As an example, suppose that the minimum acceptable reliability of a system is 0.9 ($p_1 = 0.1$) and the contractually specified reliability is 0.95 ($p_0 = 0.05$). Consumer and producer risks are to be 20%, i.e., $\alpha = \beta = 0.20$. To use the Poisson approximation, we define $\theta_0 = 1/p_0 = 1/0.05 = 20$ and $\theta_1 = 1/p_1 = 1/0.1 = 10$. The discrimination ratio, θ_0/θ_1, is 2.

TEST DESIGN FOR CONTINUOUS TIME TESTING: EXPONENTIAL MODEL

The main feature of test planning for continuously operating systems based on the exponential distribution is the assumption that the systems have a constant failure rate.

Requirement Interpretation

When the user's requirement is stated in terms of an MTBF, there is an implication of a constant failure rate. This does not mean that the system must have a constant failure rate. It means, instead, that the need remains constant. Figure E.3 illustrates that the user's needs may be met during only a portion of the time during the life of a system.

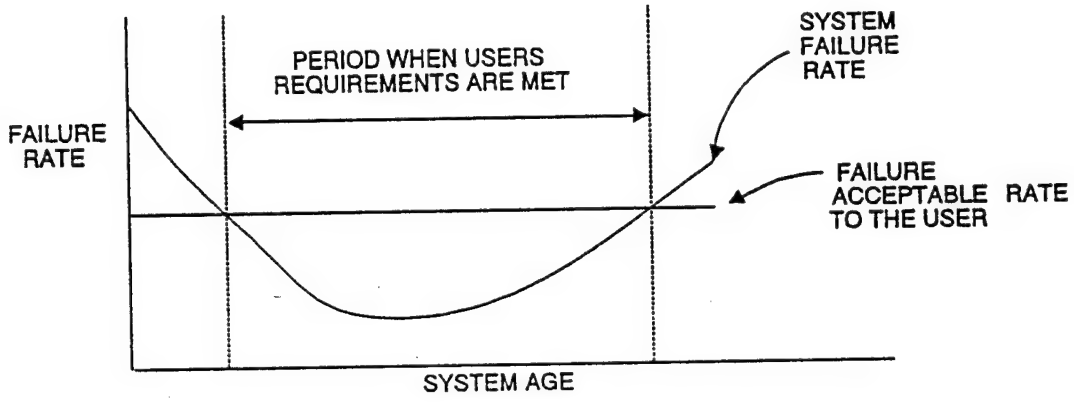


FIGURE E-3 USER REQUIREMENTS vs SYSTEM PERFORMANCE

Constant System Failure Rate Assumption

The assumption that the system to be tested has a constant failure rate may not be a good one, but it is a practical necessity for determining the amount of testing required. In theory, with the constant failure rate assumption only the total test exposure is important. That is (in theory), one system could be tested for the required test exposure, or many systems could be tested for a short time.

In practice, a test should be planned with a moderate number of systems on test for a moderate period of time. This makes the test relatively insensitive to the constant failure rate assumption. For example, one organization recommends that at least three systems be tested for at least three times the MAV (each). These are constraints imbedded in the required total test exposure.

Discrimination Ratio

The discrimination ratio, $d = \theta_0/\theta_1$, is a parameter useful in test planning for the exponential model. (For the binomial model, it is necessary to consider the upper and lower test values, p_0 and p_1, explicitly along with the α and β risks.) An interesting feature of the exponential model is that only the ratio of the upper and lower test values, $d = \theta_0/\theta_1$, along with the α and β risks need to be considered. As a consequence, test plans for the exponential models address explicitly the discrimination ratio as a planning parameter.

APPENDIX F
DATA BASE CONSIDERATIONS

INTRODUCTION

Previous appendices have presented descriptions of, and analytical techniques for, quantitatively assessing system suitability parameters - reliability, maintainability and availability. Topics discussed in these appendices included sample size, test hours, test article configuration, etc., all of which formulate the quantitative characteristics of the data base which supports our assessment.

In contrast, this appendix presents a discussion of qualitative data base characteristics. For example, it is important to conduct sufficient testing on any prototype system, but it is essential that a production decision be supported by a data base composed of production configuration test data. Also, a data base format must be structured before any actual testing is conducted to assure that the required information is collected during testing. Finally, the availability of a meaningful reliability data base, early in production, that reflects early deployment performance can be especially valuable from both a readiness and an economic viewpoint.

These and other qualitative characteristics must be considered on an a priori basis to assure that the data base under development can support the required assessment.

TEST EXPOSURE CONSIDERATIONS

Perhaps one of the most important subjects to be considered in the evaluation of RAM characteristics is the subject of test exposure. The term "test exposure" refers to the amount (quantity and quality) of testing performed on a system or systems in an effort to evaluate performance factors. The connotation of the term test exposure should include much more than what is meant by the classical "sample size." When considering single shot devices, test exposure refers to the number of items expended. On the other hand, for non-repairable, continuous operation systems, i.e., destructive testing, test exposure refers to the amount of time consumed during the test. In this situation, the number of items required for testing is not known until the test is completed, i.e., the required amount of time on test has been achieved. This results because the actual operating life of each unit is unknown until after the test is completed.

Now consider the case of non-destructive testing on single shot or continuous operation systems. For a single shot system, non-destructive test exposure refers to the number of operating cycles. All cycles could, in theory, be performed on a single item. For a continuous operation system, non-destructive test exposure refers to the amount of test time to be accumulated just as for the destructive testing case. However, with non-destructive testing, the test designer should exercise

good judgment in precisely defining the test exposure. Elements to consider are:

- Should the time required be accumulated on one system or several systems?
- Should prototypes or production models be used?
- Should testing be accelerated by eliminating nonoperating time? If so, what is the effect?
- Do we anticipate changes in equipment failure rate due to age effects or design modifications?
- Is the external environment commensurate with the requirements?

One System Vs. Several Systems

Testing one item for 100 hours in, practically speaking, not the same as testing ten items for 10 hours each, although when considering the exponential model, the two tests are theoretically equivalent. The major statistical assumptions involved are a constant failure rate and a homogeneous population. Although the two test alternatives presented are not equivalent for practical evaluation purposes, it is not a case of one being "better" than the other. Each alternative has a desirable feature. The test involving ten items has the advantage of using a greater cross-section of the population. This is particularly important if the population quality is inconsistent. On the other hand, the test of one item for 100 hours has the advantage of exploring more fully the effects of equipment age on system reliability.

As a general rule, for evaluation purposes, it is desirable to test a "moderate" number of items for a "moderate" period. This makes the test relatively insensitive to the underlying statistical assumptions of constant failure rate and sample homogeneity. One compromise between sample size and test exposure requires that a minimum of three items will operate for at least 1.5 times the minimum acceptable value (MAV). As another example, MIL-STD 781C recommends for production acceptance testing that 10% of the lot be tested, down to a minimum of 3 times, and up to a maximum of 20 items.

Another recommendation presented in MIL-STD 781C is for each test article to operate at least one half the average operating time of all articles on test. If some of the test articles experience an excessive number of failures there is a natural tendency for them to accumulate little test exposure, simply because of the difficulty of keeping them on test. A constraint of this type should minimize this biasing tendency.

Accelerated Testing and External Environment

Because the operating life of most systems generally exceeds the available test period length, some form of accelerated testing is often performed. The acceleration may consist of merely eliminating standby time from the duty cycle to subjecting the equipment to some sort of overstress conditions. The evaluator should be aware of the impact of accelerated testing on the equipment and how it will influence his analysis.

Circumstances not defined in the mission scenario can significantly impact the results of a reliability test program. Every effort should be made to control these and other effects so that the test environment is commensurate with the intended operational environment.

Prototypes Vs. Production Models

In most cases, there is no choice in this matter. For instance, in development testing there are generally only prototype models available. However, for operational testing and evaluation, production models should be used. In this case, we are trying to evaluate the "final" configuration system as it will actually perform "in the field," rather than evaluating the system's inherent capabilities.

COMPOSING THE RELIABILITY DATA BASE

When we use data analysis techniques that consider the possibility of a changing failure rate, we are acknowledging that there is reason to suspect that the failure rate may not be constant. Two of the most common causes for a changing failure rate are:

- Inherent changes in the equipment as it accumulates more hours of operation, i.e., as it ages.
- Changes in the equipment due to design changes.

There is, at present, no readily usable statistical technique for analyzing system reliability which is affected by two or more of these factors. We cannot define a precise method for evaluating test results which are derived from systems which are improving as a consequence of design modifications and at the same time degrading as a consequence of wear-out. The only guidance for this situation is to perform individual analyses on each of the subsystems for that period of time when they have a fixed configuration. Total system reliability can be obtained by piecing together the subsystem reliabilities in accordance with a system reliability model (series, parallel, etc.).

Age-Dependent Analysis

When the configurations of the systems on test are the same and fixed, we may be interested in observing the effects of aging on the failure characteristics. For this situation, we are required to record the actual age of the system when it has failed. Each element of the data base represents the age of the failed system at the time of failure.

As an example, suppose that 3 systems have been on test. System 1 operates from 0 to 1000 hours and failed 3 times. The times of failure were 20 hours, 90 hours, and 615 hours. System 2 operated from 100 to 800 hours and failed 4 times. The times of failure were 130 hours, 195 hours, 345 hours, and 520 hours. System 3 operated from 500 to 1000 hours and failed 2 times. The times of failure were 560 hours and 820 hours. The recorded times represent the age of the system at the time of failure. For an age-dependent analysis, the cumulative operating times are actually irrelevant. The data base for this test is the set of failure times.

{20, 90, 130, 195, 345, 520, 560, 615, 820}

See Figure F-1 for a graphical portrayal of composing this type of data base.

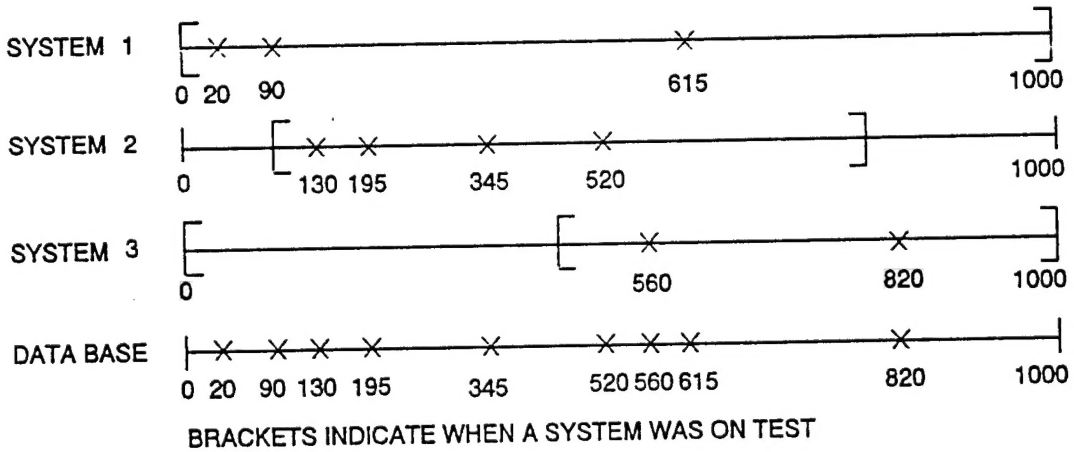


FIGURE F-1 SYSTEM FAILURE TIMES

Growth Analysis

When a system or systems are undergoing a development type test during which design modifications are being incorporated, we may be interested in observing the effects of these design changes on the reliability of the system. In this situation, the systems are being tested so

that weaknesses in design will surface as failures. Ideally, when a failure occurs, all testing will stop while the failure is analyzed and a design modification is developed. The modification is incorporated on all test systems and testing is resumed until the next failure occurs. Theoretically, for this type of testing, we are not interested in the age of each of the systems. Rather, we are interested in the cumulative time they have been on test when a failure occurs. Each element of the data base for a reliability growth analysis represents the total test time accumulated by all systems at the exact time a failure occurs.

As an example, suppose that 3 systems have been engaged in development testing. In Figure F-2 we display the failure patterns of the systems.

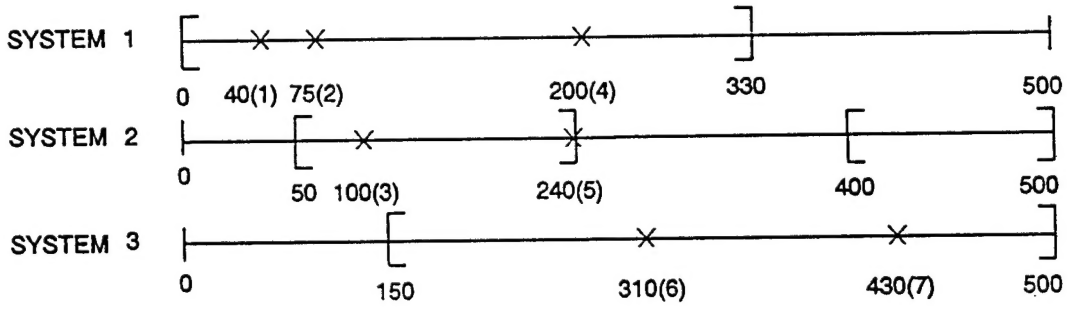


FIGURE F-2 SYSTEM TIME ON TEST

System 1 was the only one on test for the first 50 hours. System 2 began operating at the 50-hour point and System 3 at the 150-hour point. System 2 was taken off test at the 240-hour point and returned to testing at the 400-hour point. System 1 was taken off test at the 330-hour point and did not return. To compose the data base for this test we must determine how much test operating time has been accumulated when a failure has occurred.

TABLE F-1. CUMULATIVE TEST EXPOSURE/FAILURE HISTORY

Failure Number	System 1	System 2	System 3	Total Test Exposure
1	40	0	0	40
2	75	25	0	100
3	100	50	0	150
4	200	150	50	400
5	240	190	90	520
6	310	190	160	660
7	330	220	280	830

The data base for this test is the set of failure times:

{40, 100, 150, 400, 520, 660, 830}.

Note that these failure times have noting to do with the system ages. In fact, the data base is the same whether the systems have no operating time on them at the start of the test or the systems have substantial amounts of operating time at the start of the test.